

?ds

Set	Items	Description
S1	2933313	DUPLICAT? OR DUPE? ? OR COMMON? OR SAME? OR REDUNDAN? OR R- EPLICA? OR REPEAT? OR IDENTICAL? OR DOUBLE? OR ANALOGOUS? OR - SECOND OR PARALLEL? OR MULTIPLE?
S2	3685657	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?.
S3	3688754	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?
S4	3789485	REMOV? OR DELETE? OR DELETING? OR ERASE? ? OR ERASING? OR - DEDUPE? ? OR CHECK? OR FIND???? OR IDENTIF? OR DETECT? OR C- OMPAR???? OR MATCH? OR PREVENT?
S5	81570	S1(3N)S3
S6	4991	S5(5N)S4
S7	2876186	MERGE? OR MERGING? OR CONSOLIDAT? OR COMBIN? OR JOIN??? OR GROUP? OR COMPRESS? OR COMPOSITE?
S8	3357	S7(5N)S5
S9	250	S6 AND S8
S10	980056	MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL? OR CLINICAL?
S11	30	S10 AND S9
S12	29	RD (unique items)
S13	27	S12 NOT PY>2001
S14	38	S9 AND (DATABASE? OR DATA()BASE? OR DATA(2N) (WAREHOUS? OR - WARE()HOUS? OR MINE? ? OR MINING?) OR DATAMIN? OR DB OR DBS OR RECORD?()MANAGE?)
S15	37	RD (unique items)
S16	33	S15 NOT PY>2001
S17	29	S16 NOT S13

?show files

File 35:Dissertation Abs Online 1861-2004/Feb
(c) 2004 ProQuest Info&Learning

File 583:Gale Group Globalbase(TM) 1986-2002/Dec 13
(c) 2002 The Gale Group

File 65:Inside Conferences 1993-2004/Mar W4
(c) 2004 BLDSC all rts. reserv.

File 2:INSPEC 1969-2004/Mar W3
(c) 2004 Institution of Electrical Engineers

File 233:Internet & Personal Comp. Abs. 1981-2003/Sep
(c) 2003 EBSCO Pub.

File 474:New York Times Abs 1969-2004/Mar 29
(c) 2004 The New York Times

File 475:Wall Street Journal Abs 1973-2004/Mar 29
(c) 2004 The New York Times

File 99:Wilson Appl. Sci & Tech Abs 1983-2004/Feb
(c) 2004 The HW Wilson Co.

File 202:Info. Sci. & Tech. Abs. 1966-2004/Feb 27
(c) 2004 EBSCO Publishing

?

13/5/17 (Item 1 from file: 2)
DIALOG(R)File 2:INSPEC
(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6994728 INSPEC Abstract Number: C2001-09-7330-203

Title: LIFECODE: a deployed application for automated medical coding
Author(s): Heinze, D.T.; Morsch, M.L.; Sheffer, R.E.; Jimmink, M.A.; Jennings, M.A.; Morris, W.C.; Morsch, A.E.W.
Journal: AI Magazine vol.22, no.2 p.76-88
Publisher: American Assoc. Artificial Intelligence,
Publication Date: Summer 2001 **Country of Publication:** USA
CODEN: AIMA EK **ISSN:** 0738-4602
SICI: 0738-4602(200122)22:2L:76:LDAA;1-X
Material Identity Number: H885-2001-003
U.S. Copyright Clearance Center Code: 0738-4602/2001/\$2.00
Language: English **Document Type:** Journal Paper (JP)
Treatment: Practical (P)

Abstract: LIFECODE is a natural language processing (NLP) and expert system that extracts demographic and **clinical** information from free-text **clinical** records. The initial application of LIFECODE is for the emergency medicine **clinical** specialty. An application for diagnostic radiology went into production in October 2000. The LIFECODE NLP engine uses a large number of specialist readers whose particular output are combined at various levels to form an integrated picture of the **patient's** **medical** condition(s), course of treatment, and disposition. The LIFECODE expert system performs the tasks of **combining** complementary information, **deleting** redundant information, assessing the level of **medical** risk and level of service represented in the **clinical** record, and producing an output that is appropriate for input to an electronic **medical** record (EMR) system or a **hospital** information system. Because of the critical nature of the tasks, LIFECODE has a unique "self-awareness" feature that enables it to recognize the limits of its competence and, thus, ask for assistance from a human expert when faced with information that is beyond the bounds of its competence. The LIFECODE NLP and expert systems reside in various delivery packages, including online transaction processing, a web browser interface, and an automated speech recognition (ASR) interface. (26 Refs)

Subfile: C

Descriptors: **medical** expert systems; **medical** information systems; natural languages

Identifiers: LIFECODE; natural language processing; expert system; emergency medicine; diagnostic radiology; free-text **clinical** records; LIFECODE NLP engine; LIFECODE expert system; electronic **medical** record; **hospital** information system.; online transaction processing; web browser interface; automated speech recognition

Class Codes: C7330 (Biology and medical computing); C6180N (Natural language processing); C6170 (Expert systems and other AI software and techniques); C7140 (Medical administration)

Copyright 2001, IEE

13/5/19 (Item 3 from file: 2)
DIALOG(R)File 2:INSPEC
(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6911594 INSPEC Abstract Number: C2001-06-7330-047

Title: LifeCode/sup TM/-a natural language processing system for medical coding and data mining
Author(s): Heinze, D.T.; Morsch, M.L.; Sheffer, R.E., Jr.; Jimmink, M.A.; Jennings, M.A.; Morris, W.C.; Morsch, A.E.W.
Author Affiliation: A-Life Med. Inc., San Diego, CA, USA
Conference Title: Proceedings Seventeenth National Conference on Artificial Intelligence (AAAI-2000). Twelfth Innovative Applications of Artificial Intelligence Conference (IAAI-2000) p.965-72
Publisher: AAAI Press, Menlo Park, CA, USA

Publication Date: 2000 Country of Publication: USA x+1190 pp.
ISBN: 0 262 51112 6 Material Identity Number: XX-2000-02069
Conference Title: Proceedings of the Seventeenth National Conference on
Artificial Intelligence
Conference Sponsor: American Assoc. Artificial Intelligence
Conference Date: 30 July-3 Aug. 2000 Conference Location: Austin, TX,
USA

Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P)

Abstract: LifeCode/sup TM/ is a natural language processing and expert system that extracts demographic and **clinical** information from free-text **clinical** records. The initial application of LifeCode is for the emergency medicine **clinical** specialty. An application for diagnostic radiology is now in beta-test. A pilot program for performing data mining on acute care **clinical** records has been completed. The LifeCode NLP engine uses a large number of specialist readers whose particular outputs are combined at various levels to form an integrated picture of the **patient**'s **medical** condition(s), course of treatment and disposition. The LifeCode expert system performs the tasks of **combining** complementary **information**, **deleting redundant information**, assessing the level of **medical** risk and level of service represented in the **clinical** record and producing an output that is appropriate for input to an electronic **medical** record (EMR) system or a billing system. Because of the critical nature of the tasks, LifeCode has a unique "self-awareness" feature that enables it to recognize the limits of its competence and thus ask for assistance from a human expert when faced with information that is beyond the bounds of its competence. The LifeCode NLP and expert systems are wrapped as DCOM servers and reside in various delivery packages including On-Line Transaction Processing (OLTP), a Web-browser interface and an automated speech recognition (ASR) interface. (24 Refs)

Subfile: C

Descriptors: data mining; **medical** administrative data processing; **medical** diagnostic computing; **medical** expert systems; natural language interfaces

Identifiers: natural language processing system; data mining; expert system; demographic information; free-text **clinical** records; diagnostic radiology; LifeCode NLP engine; **medical** condition; treatment course; **patient** disposition; complementary information; redundant information; electronic **medical** record system; EMR system; billing system; self-awareness feature; DCOM servers; delivery packages; On-Line Transaction Processing; OLTP; online transaction processing; Web-browser interface; automated speech recognition interface; ASR) interface

Class Codes: C7330 (Biology and medical computing); C7140 (Medical administration); C6170 (Expert systems and other AI software and techniques); C6180N (Natural language processing); C6130 (Data handling techniques); C6170K (Knowledge engineering techniques); C6160 (Database management systems (DBMS))

Copyright 2001, IEE

13/5/20 (Item 4 from file: 2)
DIALOG(R)File 2:INSPEC
(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6396305 INSPEC Abstract Number: C1999-12-7140-022

Title: Intelligent data clean-up in epidemiological registers

Author(s): Hinrichs, H.

Author Affiliation: Oldenburger Forschungs- und Entwicklungsinstit. für Inf.-Werkzeuge und -Syst., Germany

Conference Title: Datenbanksysteme in Büro, Technik und Wissenschaft (Database Systems in Office, Technology and Knowledge Work) p.291-306

Editor(s): Buchmann, A.P.

Publisher: Springer-Verlag, Berlin, Germany

Publication Date: 1999 Country of Publication: Germany x+465 pp.

ISBN: 3 540 65606 5 Material Identity Number: XX-1999-00502

Conference Title: Proceedings of Datenbanksysteme in Büro Technik und Wissenschaft. BTW '99. (Database systems in the office technology and

research)

Conference Date: 1-3 March 1999 Conference Location: Freiburg im Breisgau, Germany

Language: German Document Type: Conference Paper (PA)

Treatment: Practical (P)

Abstract: Notes the heterogeneous character of **medical** epidemiological data, so that quality has to be improved before processing. Use of data **compression**, and **removal** of **redundant** data, are considered. Announces the "CARELIS" computer program (Cancer Registry Lower Saxony Record Linkage System) which integrates probabilistic record linkage and knowledge processing in order to reduce the amount of manual pre-processing required. The steps in cleaning-up data in cancer information are described and record standardisation and record linkage are explained. A block diagram of the program is presented and knowledge based message processing is considered. Applications where data is derived from a data warehouse are discussed. (17 Refs)

Subfile: C

Descriptors: cancer; data compression; data warehouses; knowledge based systems; **medical** information systems; records management

Identifiers: intelligent data clean-up; epidemiological registers; heterogeneous character; **medical** epidemiological data; data compression; redundant data; CARELIS computer program; Cancer Registry Lower Saxony Record Linkage System; probabilistic record linkage; knowledge processing; cancer information; record standardisation; knowledge based message processing; data warehouse

Class Codes: C7140 (Medical administration); C6170 (Expert systems and other AI software and techniques); C6130 (Data handling techniques); C6160Z (Other DBMS)

Copyright 1999, IEE

13/5/21 (Item 5 from file: 2)

DIALOG(R) File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6049745 INSPEC Abstract Number: C9811-7140-106

Title: Combining **laboratory** data sets from multiple institutions using the Logical Observation Identifier Names and Codes (LOINC)

Author(s): Baorto, D.M.; Cimino, J.J.; Parvin, C.A.; Kahn, M.G.

Author Affiliation: Dept. of Pathology, Washington Univ. Sch. of Med., St. Louis, MO, USA

Journal: International Journal of Medical Informatics vol.51, no.1 p.29-37

Publisher: Elsevier,

Publication Date: July 1998 Country of Publication: Ireland

CODEN: IJMIF4 ISSN: 1386-5056

SICI: 1386-5056(199807)51:1L:29:CLDS;1-A

Material Identity Number: G133-98008

U.S. Copyright Clearance Center Code: 1386-5056/98/\$19.00

Document Number: S1386-5056(98)00089-6

Language: English Document Type: Journal Paper (JP)

Treatment: Experimental (X)

Abstract: A standard set of names and codes for laboratory test results is critical for any endeavor requiring automated data pooling, including multi-institutional research and cross-facility **patient** care. This need has led to the development of the Logical Observation Identifier Names and Codes (LOINC) database and its test-naming convention. For this study, we required two **medical** centers to extract raw laboratory data from their local information system for a defined **patient** population, translate tests into LOINC and provide aggregate data which could then be used to compare laboratory utilization. We examine a larger number of tests from each site which have been recoded using an updated version of the LOINC database. We conclude that the coding of local tests into LOINC can often be complex, especially the 'Kind of Property' field, and apparently trivial differences in choices made by individual institutions can result in nonmatches in electronically pooled data. 75% of failures to match the same tests between different institutions using LOINC codes were due to

differences in local coding choices. LOINC has the potential to eliminate the need for detailed human inspection during the pooling of laboratory data from diverse sites and perhaps even a built-in capability to adjust matching stringency by selecting subsets of LOINC fields required to match. However, a quality standard coding procedure is required, and examples highlighted in this paper may require special attention while mapping to LOINC. (12 Refs)

Subfile: C

Descriptors: codes; database management systems; **medical** information systems; merging; nomenclature

Identifiers: laboratory data sets; Logical Observation Identifier Names and Codes; LOINC database; laboratory test results; automated data pooling; multi-institutional research; cross-facility **patient** care; **medical** centers; raw laboratory data extraction; local information system; **patient** population; laboratory utilization; property type; local coding choices; detailed human inspection; matching stringency; database fields; quality standard coding procedure

Class Codes: C7140 (Medical administration); C6130 (Data handling techniques); C6160Z (Other DBMS)

Copyright 1998, IEE

13/5/22 (Item 6 from file: 2)

DIALOG(R) File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

5787533 INSPEC Abstract Number: B9802-7540-002, C9802-7140-082

Title: Medical **images** and electronic **patient** **information:** finding the link

Author(s): Goel, H.

Author Affiliation: IMNET Syst. Inc., USA

Conference Title: Toward an Electronic Patient '97. Conference and Exposition. Proceedings Part vol.2 p.234-5 vol.2

Editor(s): Waegemann, C.P.

Publisher: Med. Records Inst, Newton, MA, USA

Publication Date: 1997 **Country of Publication:** USA 3 vol. (387+324+379) pp.

ISBN: 0 9640667 9 3 **Material Identity Number:** XX97-03129

Conference Title: Proceedings of TEPR '97. Toward an Electronic Patient Record '97

Conference Date: 27 April-3 May 1997 **Conference Location:** Nashville, TN, USA

Language: English **Document Type:** Conference Paper (PA)

Treatment: General, Review (G)

Abstract: **Medical** images and electronic **patient** information have been separated by a seemingly impenetrable barrier. With the ever-growing influence of managed care, the race is now on to electronically link these vital sources of information. Traditional PACS solutions have come under increasing fire due to their proprietary nature and extremely high cost.

Health care providers are all racing to establish a computerized **patient** record (CPR), and they realize that **medical** images are an integral part of the CPR. They also realize that the disparate nature of current PACS solutions will not allow them to integrate **medical** images with the rest of their **patient** data. A link between the two must be found to make the CPR a reality. The solution is an integration between the enterprise's **health** care information system (HIS) and on-line **medical** images; an infoPACS solution. InfoPACS are critical to the information integration plans of today's **health** care providers. They supply the missing link between radiology and the rest of the enterprise by providing PACS image acquisition, display, and archiving components, and integration between **medical** images and electronic **patient** information. This allows users in radiology and throughout the enterprise to view every aspect of a **patient**'s **medical** record at any time, from anywhere. By **combining** all **patient** **information** together in the **same** database, infoPACS solutions **remove** a major barrier to true enterprise-wide integration. (0 Refs)

Subfile: B C

Descriptors: integrated software; **medical** information systems; PACS

Identifiers: **medical** images; electronic **patient** information; managed care; PACS; computerized **patient** record; **health** care information system; infoPACS; information integration; PACS image acquisition; archiving; **medical** record; **patient** information; enterprise-wide integration
Class Codes: B7540 (Hospital Engineering); B7510B (Radiation and radioactivity applications in biomedicine); B6210L (Computer communications); C7140 (Medical administration); C5260B (Computer vision and image processing techniques); C5620L (Local area networks); C7330 (Biology and medical computing)

Copyright 1997, IEE

13/5/26 (Item 10 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

01134925 INSPEC Abstract Number: C78001815

Title: An experiment to measure the performance of phonetic key compression retrieval schemes

Author(s): Greenfield, R.H.

Author Affiliation: Washington Univ. School of Medicine, St. Louis, MO, USA

Journal: Methods of Information in Medicine vol.16, no.4 p.230-3

Publication Date: Oct. 1977 Country of Publication: West Germany

CODEN: MIMCAI ISSN: 0026-1270

Language: English Document Type: Journal Paper (JP)

Treatment: Experimental (X)

Abstract: The results of an experiment to measure the performance of the Davidson and Soundex phonetic key **compression** schemes in **finding** sets of **records** representing the **same** individual in a moderately large radiology **patient** file as compared to exact surname matches, are presented. This is similar to the problem of retrieving a record by name under the assumption that neither the search key nor the recorded key is accurately known. Both phonetic schemes perform similarly in obtaining extra matches and both outperform by a large margin the exact name match. The results also indicate that the Davidson scheme is superior to the Soundex because it produces significantly fewer mismatches. (7 Refs)

Subfile: C

Descriptors: data handling; **medical** administrative data processing

Identifiers: experiment; measure; performance; phonetic key compression; retrieval schemes; Davidson; Soundex; records; radiology **patient** file

Class Codes: C6130 (Data handling techniques); C7140 (Medical administration)

13/5/27 (Item 1 from file: 202)

DIALOG(R)File 202:Info. Sci. & Tech. Abs.

(c) 2004 EBSCO Publishing. All rts. reserv.

1301411

An experiment to measure the performance of phonetic key compression retrieval schemes.

Author(s): Greenfield, R H

Corporate Source: Washington University School Of Medicine, St. Louis, Missouri.

Methods of Information in Medicine vol. 16, no. 4, pages 230-233

Publication Date: 1977

ISSN: 0026-1270

Language: English

Document Type: Journal Article

Record Type: Abstract

Journal Announcement: 1300

The results of an experiment to measure the performance of the davidson and soundex phonetic key **compression** schemes in **finding** sets of **records** representing the **same** individual in a moderately large radiology **patient** file, as compared to exact surname matches, are presented. This

is similar to the problem of retrieving a record by name under the assumption that neither the search key nor the recorded key is accurately known. Both phonetic schemes perform similarly in obtaining extra matches and both outperform by a large margin the exact name match. The results also indicate that the davidson scheme is superior to the soundex because it produces significantly fewer mismatches.

Classification Codes and Description: 5.11 (Searching and Retrieval)

Main Heading: Information Processing and Control

?

17/5/1 (Item 1 from file: 35)
DIALOG(R)File 35:Dissertation Abs Online
(c) 2004 ProQuest Info&Learning. All rts. reserv.

01603431 ORDER NO: AAD98-04033
**ADAPTIVE DETECTION OF APPROXIMATELY DUPLICATE DATABASE RECORDS AND
THE DATABASE INTEGRATION APPROACH TO INFORMATION DISCOVERY**
Author: MONGE, ALVARO EDMUNDO
Degree: PH.D.
Year: 1997
Corporate Source/Institution: UNIVERSITY OF CALIFORNIA, SAN DIEGO (0033)
Chairperson: CHARLES P. ELKAN
Source: VOLUME 58/08-B OF DISSERTATION ABSTRACTS INTERNATIONAL.
PAGE 4315. 98 PAGES
Descriptors: COMPUTER SCIENCE ; INFORMATION SCIENCE
Descriptor Codes: 0984; 0723

The integration of information is an important area of research in **databases**. By **combining multiple information** sources, a more complete view of the world is attained, and additional knowledge gained. This is a non-trivial task however. Often there are many sources which contain information about a certain kind of entity, and some will contain records concerning the same real-world entity. Thus, one problem in integrating information sources is to identify possibly different designators of the same entity. This thesis provides solutions to this data cleansing problem. The integration of information sources is also proposed as an approach for information retrieval over the worldwide web.

Data cleansing is the process of purging **databases** of inaccurate or inconsistent data. The data is manipulated into a form which is useful for other tasks, such as **data mining**. This thesis addresses the data cleansing problem of **detecting database records** that are approximate **duplicates**, but not exact duplicates. An efficient algorithm is presented which combines three key ideas. First, the Smith-Waterman algorithm for computing the minimum edit-distance is used as a domain-independent method to recognize pairs of approximately duplicates. **Second**, the union- **find data** structure is used to maintain the clusters of duplicate records incrementally, as pair-wise duplicate relationships are discovered. Third, the algorithm uses a priority queue of cluster subsets to respond adaptively to the size and homogeneity of the clusters discovered as the **database** is scanned. This results in significant savings in the number of times that a pairwise record matching algorithm is applied, without impairing accuracy. Comprehensive experiments on synthetic **databases** and on a real-world **database** confirm the effectiveness of all three ideas.

This thesis also presents W scEBF scIND, an application that discovers scientific papers online over the worldwide web. The size of the web makes online information retrieval difficult. In such a setting, it is critical to know where to concentrate the search for information. W scEBF scIND uses a domain-independent algorithm to match records from different sources for the purpose of integrating the information in those sources. We describe the design of W scEBF scIND, the integration process, and discovery phase.

17/5/5 (Item 1 from file: 2)
DIALOG(R)File 2:INSPEC
(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6796252 INSPEC Abstract Number: C2001-02-7330-065

Title: Mouse genome informatics in a new age of biological inquiry
Author(s): Bult, C.J.; Richardson, J.E.; Blake, J.A.; Kadin, J.A.; Ringwald, M.; Eppig, J.T.; Baldarelli, R.M.; Baya, M.; Beal, J.S.; Begley, D.A.; Boddy, W.J.; Bradt, D.W.; Butler, N.E.; Chu, T.; Corbani, L.E.; Corradi, J.; Davisson, M.T.; Garippa, D.; Glass, L.H.; Grant, P.L.; Hill, D.P.; King, B.L.; Krupke, D.M.; Lennon-Pierce, M.; Lutz, C.M.; Maltais, L.J.; Mani, P.; McCright, I.; McKenzie, L.; Naf, D.; Ormsby, J.E.; Ramachandran, S.; Reed, D.J.; Shaw, D.R.; Szauter, P.; Trombley, L.A.

Author Affiliation: Ja [redacted] on Lab., Bar Harbor, ME, USA [redacted]
Conference Title: Proceedings IEEE International Symposium on
Bio-Informatics and Biomedical Engineering p.29-32
Publisher: IEEE Comput. Soc, Los Alamitos, CA, USA
Publication Date: 2000 Country of Publication: USA xi+386 pp.
ISBN: 0 7695 0862 6 Material Identity Number: XX-2000-02862
U.S. Copyright Clearance Center Code: 0 7695 0862 6/2000/\$10.00
Conference Title: Proceedings IEEE International Symposium on
Bio-Informatics and Biomedical Engineering
Conference Sponsor: IEEE Comput. Soc.; IEEE Comput. Soc. Task Force on
Virtual Intelligence (TFVI); IEEE Eng. Med. & Biology Soc.; EMB; AIIS;
ICIIS; ICTAI; NIH; NCI; TIGR
Conference Date: 8-10 Nov. 2000 Conference Location: Arlington, VA,
USA

Language: English Document Type: Conference Paper (PA)
Treatment: Practical (P)

Abstract: Data integration is key to knowledge discovery in the age of
genomics and represents a major, long-standing challenge for the genome
informatics community. Integration of data across heterogeneous genome
databases requires the **identification** of **common data** entities and
mechanisms to ensure referential integrity and persistence of these common
entities even as our understanding of their biological properties changes.
The Mouse Genome Informatics (MGI) **database group** employs **multiple**
strategies for achieving **data** integration and maintaining
inter-connectedness with other **databases**, including: (1) the use of
permanent, unique accession IDs for identifying core data entities, (2) the
application of nomenclature standards for naming genes and strains of mice,
and (3) the development and implementation of controlled vocabularies and
ontologies to ensure semantic consistency of biological concepts within and
across model organism **databases**. (10 Refs)

Subfile: C

Descriptors: biology computing; data integrity; **data mining** ;
distributed **databases** ; genetics; nomenclature; scientific information
systems; standards; vocabulary; zoology

Identifiers: mouse genome informatics; biological inquiry; data
integration; knowledge discovery; genomics; heterogeneous genome **databases**
; common data entities; referential integrity; biological properties;
database interconnectedness; accession IDs; core data entities;
nomenclature standards; gene naming; mouse strains; controlled vocabularies
; ontologies; semantic consistency; biological concepts; model organism
databases

Class Codes: C7330 (Biology and medical computing); C6160B (Distributed
databases); C7240 (Information analysis and indexing)

Copyright 2000, IEE

17/5/6 (Item 2 from file: 2)
DIALOG(R) File 2:INSPEC
(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6552308 INSPEC Abstract Number: C2000-05-6160S-025

Title: Data redundancy and duplicate detection in spatial join
processing

Author(s): Dittrich, J.-P.; Seeger, B.

Author Affiliation: Dept. of Math. & Comput. Sci., Marburg Univ., Germany

Conference Title: Proceedings of 16th International Conference on Data
Engineering (Cat. No.00CB37073) p.535-46

Publisher: IEEE Comput. Soc, Los Alamitos, CA, USA

Publication Date: 2000 **Country of Publication:** USA xxii+703 pp.

ISBN: 0 7695 0506 6 **Material Identity Number:** XX-2000-00609

U.S. Copyright Clearance Center Code: 0 7695 0506 6/2000/\$10.00

Conference Title: Proceedings 16th International Conference on Data
Engineering

Conference Sponsor: IEEE Comput. Soc. Tech. Committee on Data Eng

Conference Date: 29 Feb.-3 March 2000 **Conference Location:** San Diego,
CA, USA

Language: English **Document Type:** Conference Paper (PA)

Treatment: Practical () Experimental (X)

Abstract: The partition-based spatial-merge join (PBSM) of J.M. Patel and D.J. DeWitt (1996) and the size separation spatial join (S/sup 3/J) of N. Koudas and K.C. Sevcik (1997) are considered to be among the most efficient methods for processing spatial (intersection) joins on two or more spatial relations. Neither method assumes the presence of pre-existing spatial indices on the relations. In this paper, we propose several improvements to these join algorithms. In particular, we deal with the impact of **data redundancy** and **duplicate detection** on the performance of these methods. For PBSM, we present a simple and inexpensive online method to detect duplicates in the response set. There is no longer any need to eliminate duplicates in a final sorting phase, as was originally suggested. We also investigate the impact of different internal algorithms on the total run-time of PBSM. For S/sup 3/J, we break with the original design goal and introduce controlled redundancy of data objects. Results of a large set of experiments with real data sets reveal that our suggested modifications to PBSM and S/sup 3/J result in substantial performance improvements, where PBSM is generally superior to S/sup 3/J. (33 Refs)

Subfile: C

Descriptors: data reduction; merging; query processing; redundancy; software performance evaluation; visual **databases**

Identifiers: data redundancy; duplicate detection; spatial join processing; partition-based spatial-merge join; size separation spatial join; intersection joins; spatial relations; performance; online method; response set; sorting phase; internal algorithms; run-time; controlled data object redundancy

Class Codes: C6160S (Spatial and pictorial databases); C6130 (Data handling techniques)

Copyright 2000, IEE

17/5/9 (Item 5 from file: 2)

DIALOG(R) File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6308664 INSPEC Abstract Number: C1999-09-6160D-008

Title: Merge replication in Microsoft's SQL Server 7.0

Author(s): Hammond, B.

Journal: SIGMOD Record Conference Title: SIGMOD Rec. (USA) vol.28, no.2 p.527

Publisher: ACM,

Publication Date: June 1999 Country of Publication: USA

CODEN: SRECD8 ISSN: 0163-5808

SICI: 0163-5808(199906)28:2L:527:MRMS;1-O

Material Identity Number: A660-1999-002

U.S. Copyright Clearance Center Code: 0163-5808/99/\$05...\$5.00

Conference Title: 1999 ACM SIGMOD International Conference on Management of Data

Conference Date: 1-3 June 1999 Conference Location: Philadelphia, PA, USA

Language: English Document Type: Conference Paper (PA); Journal Paper (JP)

Treatment: Practical (P)

Abstract: SQL Server 7.0 offers three different styles of replication: transactional replication, snapshot replication, and merge replication. The author focuses on **merge replication**. This means that **data** changes can be performed at any replica, and that the changes performed at multiple replicas are later merged together. Because merge replication allows updates to disconnected replicas, it is particularly well suited to applications that require a lot of autonomy. A special process called the **merge** agent propagates changes between **replicas**, filters **data** as appropriate, and **detects** and handles conflicts according to user-specified rules. (0 Refs)

Subfile: C

Descriptors: client-server systems; merging; query processing; relational **databases**; replicated **databases**; SQL

Identifiers: Microsoft; SQL Server; transactional replication; snapshot

replication; merge replication; data changes; disconnect replicas; merge agent; user-specified rules

Class Codes: C6160D (Relational databases); C6140D (High level languages)
; C6160B (Distributed databases); C6150N (Distributed systems software)
Copyright 1999, IEE

17/5/10 (Item 6 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6006014 INSPEC Abstract Number: C9810-7250C-001

Title: Duplicate detection and record consolidation in large bibliographic databases : the COPAC database experience

Author(s): Cousins, S.A.

Author Affiliation: Manchester Comput., Manchester Univ., UK

Journal: Journal of Information Science vol.24, no.4 p.231-40

Publisher: Bowker-Saur for Inst. Inf. Sci,

Publication Date: 1998 **Country of Publication:** UK

CODEN: JIOSED **ISSN:** 0165-5515

SICI: 0165-5515(1998)24:4L.231:DDRC;1-1

Material Identity Number: C262-98004

Language: English **Document Type:** Journal Paper (JP)

Treatment: Practical (P)

Abstract: COPAC is a union catalogue giving access to the online catalogue records of some of the largest academic research libraries in the UK and Ireland. Like all union catalogues COPAC is supplied with multiple copies of records representing the same document in the contributing library catalogues. To reduce the level of duplication visible to the COPAC user, it has been necessary to develop **duplicate detection** and **record consolidation** procedures. These result in the production of a single record for each document, representing the holdings of several libraries. This paper discusses the ways in which both the **duplicate detection** and **record consolidation** procedures are carried out, along with the problem areas encountered. The general structure of these procedures is also described, providing a model of the duplicate record handling mechanisms used in COPAC. (6 Refs)

Subfile: C

Descriptors: academic libraries; bibliographic systems; cataloguing;

library automation; research libraries; very large **databases**

Identifiers: duplicate detection; record consolidation; large

bibliographic **databases** ; COPAC **database** ; union catalogue; online catalogue records; academic research libraries; duplicate record handling

Class Codes: C7250C (Bibliographic retrieval systems); C7210L (Library automation); C7240 (Information analysis and indexing)

Copyright 1998, IEE

17/5/12 (Item 8 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

5511892 INSPEC Abstract Number: C9704-6160J-003

Title: An object deputy model for realization of flexible and powerful objectbases

Author(s): Kambayashi, Y.; Zhiyong Peng

Author Affiliation: Dept. of Inf. Sci., Kyoto Univ., Japan

Journal: Journal of Systems Integration Conference Title: J. Syst. Integr. (Netherlands) vol.6, no.4 p.329-62

Publisher: Kluwer Academic Publishers,

Publication Date: Nov. 1996 **Country of Publication:** Netherlands

CODEN: JSINE4 **ISSN:** 0925-4676

SICI: 0925-4676(199611)6:4L.329:ODMR;1-J

Material Identity Number: O661-96005

U.S. Copyright Clearance Center Code: 0925-4676/96/\$8.50

Conference Title: 4th International Conference on Database Systems for Advanced Applications

Conference Date: 10-13 April 1995 Conference Location: Singapore
Language: English Document Type: Conference Paper (PA); Journal Paper (JP)

Treatment: Theoretical (T)

Abstract: In object-oriented (OO) **databases**, data and methods are **combined** by objects and their classes. **Information redundancies** are **removed** by the notions of class hierarchy and inheritance. This contributes to the realization of high-performance systems. However, the flexibility of object structures is a problem, due to its encapsulation feature. We introduce an object deputy model which extends conventional OO models with the concepts of deputy objects and deputy classes. A deputy object has its own identifier and may possess its own attributes and methods. It can also have attributes that are computed from values stored within its source object, and can be associated with methods generated from these of the source object. The inheritances are realized by switching operations, which make it possible to realize controllable, selective and dynamic inheritance structures. Schemata of deputy objects are defined by deputy classes which can be derived by an object deputy algebra. An object can have many deputy objects, and several objects can share a single deputy object. Thus, objects can be indirectly divided and combined through their deputy objects. Several difficult **database** problems, such as flexible views, objects with more than one role, object migration and multiple inheritance, become much easier in this model. The data-knowledge coordination model developed for the integration of distributed **databases** and knowledge-bases can also be realized easily by the object deputy model. We discuss several advanced **database** applications: geographic **databases**, virtual office systems and distance education systems. (37 Refs)

Subfile: C

Descriptors: data encapsulation; **database** theory; deductive **databases**; distributed **databases**; inheritance; object-oriented **databases**

Identifiers: object deputy model; powerful flexible object-oriented **databases**; **information redundancy removal**; class hierarchy; flexible views; high-performance systems; object structure flexibility; encapsulation; deputy objects; deputy classes; switching operations; dynamic inheritance structures; object roles; object migration; multiple inheritance; data-knowledge coordination model; distributed **databases**; knowledge bases; geographic **databases**; virtual office systems; distance education systems

Class Codes: C6160J (Object-oriented databases); C6120 (File organisation); C4250 (Database theory)

Copyright 1997, IEE

17/5/13 (Item 9 from file: 2)

DIALOG(R) File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

5449960 INSPEC Abstract Number: C9701-6160K-035

Title: The field matching problem: algorithms and applications

Author(s): Monge, A.E.; Elkan, C.P.

Author Affiliation: Dept. of Comput. Sci. & Eng., California Univ., San Diego, La Jolla, CA, USA

Conference Title: KDD-96 Proceedings. Second International Conference on Knowledge Discovery and Data Mining p.267-70

Editor(s): Simoudis, E.; Han, J.; Fayyad, U.

Publisher: AAAI Press, Menlo Park, CA, USA

Publication Date: 1996 Country of Publication: USA xiv+391 pp.

Material Identity Number: XX96-02511

Conference Title: Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)

Conference Date: 2-4 Aug. 1996 Conference Location: Portland, OR, USA

Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P)

Abstract: To **combine** information from heterogeneous sources, equivalent **data** in the **multiple** sources must be **identified**. This task is the field matching problem. Specifically, the task is to determine whether or not two syntactic values are alternative designations of the same semantic

entity. For example, addresses "Dept. of Compu Sci. and Eng., University of California, San Diego, 9500 Gilman Dr. Dept. 0114, La Jolla, CA 92093" and "UCSD, Computer Science and Engineering Department, CA 92093-0114" designate the same department. This paper describes three field matching algorithms, and evaluates their performance on real-world datasets. One proposed method is the well-known Smith-Waterman (1981) algorithm for comparing DNA and protein sequences. Several applications of field matching in knowledge discovery are described briefly, including WEBFIND, which is a new software tool that discovers scientific papers published on the World Wide Web (WWW). WEBFIND uses external information sources to guide its search for authors and papers. Like many other WWW tools, WEBFIND needs to solve the field matching problem in order to navigate between information sources. (12 Refs)

Subfile: C

Descriptors: bibliographic systems; deductive **databases**; knowledge acquisition; merging; online front-ends; software performance evaluation; string matching

Identifiers: field matching problem; heterogeneous information sources; information combination; equivalent data identification; syntactic values; alternative designations; semantic entity; performance evaluation; DNA sequences; protein sequences; knowledge discovery; WEBFIND; scientific papers; World Wide Web; external information sources; author search; information navigation; addresses

Class Codes: C6160K (Deductive databases); C6170T (Knowledge engineering tools); C7250N (Front end systems for online searching); C7250C (Bibliographic retrieval systems)

Copyright 1996, IEE

17/5/16 (Item 12 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

03967650 INSPEC Abstract Number: C91059969

Title: Should journal references be standardized?

Author(s): Hudnut, S.K.

Conference Title: 12th National Online Meeting. Proceedings 1991 p. 149-55

Editor(s): Williams, M.E.

Publisher: Learned Inf, Medford, NJ, USA

Publication Date: 1991 Country of Publication: USA xii+458 pp.

ISBN: 0 938734 51 2

Conference Sponsor: Learned Inf

Conference Date: 7-9 May 1991 Conference Location: New York, NY, USA

Language: English Document Type: Conference Paper (PA)

Treatment: General, Review (G); Practical (P)

Abstract: References to journal articles are a very large part of online searching. Yet **identifying** which online **records** are **duplicates** of each other isn't always easy. The handling of titles, authors, journal names and source information across **databases** can be different enough to require close scrutiny before determining if a **group** of **records** refers to the **same** article. Using duplicate **detection** or scanning journal name indexes points up the many variations that can exist between **databases**. Is standardization the answer? If so, by whom? The paper examines the issues from the perspective of the three participants: the **database** producer, the online service and the user, and considers the impact of the existing problems on online searching. (0 Refs)

Subfile: C

Descriptors: information services; standardisation

Identifiers: journal references; online searching; standardization; **database** producer; user

Class Codes: C7210 (Information services and centres)

17/5/17 (Item 13 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

03750880 INSPEC Abstract Number: C90072173

Title: Detecting duplicates: a searcher's dream come true

Author(s): Miller, C.

Author Affiliation: Carmen Miller Resources, Seattle, WA, USA

Journal: Online vol.14, no.4 p.27-34

Publication Date: July 1990 **Country of Publication:** USA

CODEN: ONLIDN **ISSN:** 0146-5422

Language: English **Document Type:** Journal Paper (JP)

Treatment: Practical (P)

Abstract: DIALOG's announcement of a **duplicate record detection** feature in 1989 was met with enthusiasm. Searchers have been requesting a way to **identify duplicate records** since long before the 1987 introduction of OneSearch DIALOG's multiple-file search capability. Once one begins to search multiple **databases** with overlapping sources, the need for a way to sort **merged** output, **identify** the **duplicates** and retain the **records** one chooses becomes imperative. Duplicate **detection** is a long-awaited, major system enhancement which allows searchers to do all of these things. Duplicate detection was introduced in approximately 200 of DIALOG's bibliographic files and there are plans to add 60 to 70 more of the full-text newspaper, newswire, and journal article files. The author explains how she tested the DIALOG feature, how it works and its performance. (0 Refs)

Subfile: C

Descriptors: bibliographic systems; information services

Identifiers: **duplicate record detection** feature; OneSearch; multiple-file search capability; multiple **databases**; overlapping sources; merged output; system enhancement; searchers; bibliographic files; full-text newspaper; newswire; journal article files; DIALOG feature

Class Codes: C7250C (Bibliographic systems); C7210 (Information services and centres)

17/5/20 (Item 16 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

01490007 INSPEC Abstract Number: C80013320

Title: Automatic merging of monographic data bases - identification of duplicate records in multiple files : the IUCS scheme

Author(s): Williams, M.E.; MacLaury, K.D.

Author Affiliation: Information Retrieval Res. Lab., Univ. of Illinois, Urbana, IL, USA

Journal: Journal of Library Automation vol.12, no.2 p.156-68

Publication Date: June 1979 **Country of Publication:** USA

CODEN: JLAUAY **ISSN:** 0022-2240

Language: English **Document Type:** Journal Paper (JP)

Treatment: Practical (P)

Abstract: A research project entitled 'A State-wide Catalog Feasibility Study' was funded by the Illinois State Library and carried out within the Information Retrieval Research Laboratory (IRRL) of the Coordinated Science Laboratory (CSL) at the University of Illinois in Urbana-Champaign. The purpose of the project was to develop a machine algorithm for locating and pairing duplicate records in machine-readable bibliographic files in order to eliminate duplicates and permit files from different libraries to be used for creation of a union catalog of machine-readable bibliographic records. A prototype system called IUCS was developed and tested on sample files from OCLC, Northwestern University, and the University of Chicago. (1 Refs)

Subfile: C

Descriptors: library mechanisation

Identifiers: monographic **data bases**; duplicate records; multiple files; IUCS scheme; Information Retrieval Research Laboratory; Coordinated Science Laboratory; machine algorithm; library mechanisation

Class Codes: C7210 (Information services and centres)

17/5/21 (Item 1 from file: 233)
DIALOG(R)File 233:Internet & Personal Comp. Abs.
(c) 2003 EBSCO Pub. All rts. reserv.

00552146 99IT11-040

OCLC introduces SiteSearch 4.1

Information Today , November 1, 1999 , v16 n10 p57-58, 2 Page(s)

ISSN: 8755-6286

Company Name: OCLC

URL: <http://www.oclc.org>

Product Name: SiteSearch 4.1

Languages: English

Document Type: Product Announcement

Hardware/Software Compatibility: IBM PC Compatible

Geographic Location: United States

Announces the release of version 4.1 of the OCLC SiteSearch (\$NA) suite of software from OCLC of Dublin, OH (614). Reports that it provides a comprehensive solution for managing distributed library information resources in a Web environment. Says that it offers a toolkit that lets libraries integrate their electronic resources under one Web interface, provide flexible access to those resources, and build unique local **databases** . Notes that it includes the following features: Web-based record creation, vocabulary-assisted searching, **merged** results sets, **duplicate record detection** , more starter interface options, and online updates. Says that it uses Extensive Markup Language (XML) and Resource Description Framework (RDF) within its new record-building component. Adds that the **Database** Builder component uses RDF and XML as the underpinnings of a Java-based toolkit. (KMH)

Descriptors: Library Automation; XML; Java; Library; **Database** ;
Interface; **Records Management**

Identifiers: SiteSearch 4.1; OCLC

17/5/22 (Item 2 from file: 233)
DIALOG(R)File 233:Internet & Personal Comp. Abs.
(c) 2003 EBSCO Pub. All rts. reserv.

00540130 99IW07-112

Low-cost Double Take keeps data honest

Hollander, Geoffrey

InfoWorld , July 12, 1999 , v21 n28 p56, 1 Page(s)

ISSN: 0199-6649

Company Name: PeopleSmith Software

URL: <http://www.peoplesmith.com>

Product Name: Double Take 4.1

Languages: English

Document Type: Software Review

Grade (of Product Reviewed): B

Hardware/Software Compatibility: IBM PC Compatible

Geographic Location: United States

Presents a favorable review of Double Take 4.1 (\$3,000), a **database** cleaning and conditioning utility program from PeopleSmith Software of Scituate, MA (800). Runs on IBM PC Compatibles with Microsoft Windows NT, Windows 98 and Windows 95. Says that it is capable of handling a variety of files, file types, and sizes. Cites features such as an array of problem-solving alternatives, rapid processing time, and flexible output options. Reports, however, that it takes time for the user to master it and it does not work with SQL Server 7. Concludes that it is a good solution for **finding** and removing **duplicate records** or to **merge** unique records from disparate **databases** . Received a rating of four on a scale of one to five. Includes one product summary. (MEM)

Descriptors: Utility Program; **Database** Management; Enterprise Computing; File Management; Upgrade

Identifiers: Double Take 4.1; PeopleSmith Software

17/5/26 (Item 1 from file: 99)

DIALOG(R)File 99:Wilson Pl. Sci & Tech Abs
(c) 2004 The HW Wilson Co. All rts. reserv.

2375733 H.W. WILSON RECORD NUMBER: BAST01101731

LIFECODE A Deployed Application for Automated Medical Coding

Heinze, Daniel T; Morsch, Mark L; Sheffer, Ronald E

AI Magazine v. 22 no2 (Summer 2001) p. 76-88

DOCUMENT TYPE: Feature Article ISSN: 0738-4602 LANGUAGE: English

RECORD STATUS: Corrected or revised record

ABSTRACT: LIFECODE is a natural language processing (NLP) and expert system that extracts demographic and clinical information from free-text clinical records. The initial application of LIFECODE is for the emergency medicine clinical specialty. An application for diagnostic radiology went into production in October 2000. The LIFECODE NLP engine uses a large number of specialist readers whose particular output are combined at various levels to form an integrated picture of the patient's medical condition(s), course of treatment, and disposition. The LIFECODE expert system performs the tasks of **combining complementary information**, **deleting redundant information**, assessing the level of medical risk and level of service represented in the clinical record, and producing an output that is appropriate for input to an electronic medical record (EMR) system or a hospital information system. Because of the critical nature of the tasks, LIFECODE has a unique "self-awareness" feature that enables it to recognize the limits of its competence and, thus, ask for assistance from a human expert when faced with information that is beyond the bounds of its competence. The LIFECODE NLP and expert systems reside in various delivery packages, including online transaction processing, a web browser interface, and an automated speech recognition (ASR) interface. Reprinted by permission of the publisher.

DESCRIPTORS: Natural language processing; Medical **databases** ;

17/5/27 (Item 1 from file: 202)

DIALOG(R)File 202:Info. Sci. & Tech. Abs.

(c) 2004 EBSCO Publishing. All rts. reserv.

3700040

A knowledge-based approach for duplicate elimination in data cleaning.

Author(s): Low, Wai Lup; Lee, Mong Li; Ling, Tok Wang

Corporate Source: National University, Singapore 117543 ; National

University, Singapore 117543 ; National University, Singapore 117543

Information Systems vol. 26, no. 8, pages 585-606

Publication Date: December 2001

ISSN: 0306-4379

Publisher URL: <http://www.elsevier.nl>

Language: English

Document Type: Journal Article

Record Type: Abstract

Journal Announcement: 3701

Existing duplicate elimination methods for data-cleaning work on the basis of computing the degree of similarity between nearby records in a sorted **database**. High recall can be achieved by accepting records with low degrees of similarity as duplicates, at the cost of lower precision. High precision can be achieved analogously at the cost of lower recall. This is the recall-precision dilemma. Develops a generic knowledge-based framework for effective data cleaning that can implement any existing data cleaning strategies and more. Proposes a new method for computing transitive closure under uncertainty for dealing with the **merging of groups of inexact duplicate records**, and explains why small changes to window sizes has little effect on the results of the sorted neighborhood method. Experimental results with two real-world datasets show that this approach can accurately identify duplicates and anomalies with high recall and precision, thus effectively resolving the recall-precision dilemma.

Descriptors: Information retrieval; Precision; Recall; Duplicate
detection

Classification Codes and Description: 1.4 (Information retrieval research)

Main Heading: Information Science Research

17/5/28 (Item 2 from file: 202)

DIALOG(R) File 202:Info. Sci. & Tech. Abs.

(c) 2004 EBSCO Publishing. All rts. reserv.

3304784

Duplicate detection and record consolidation in large bibliographic
databases : the COPAC database experience.

Author(s): Cousins, Shirley Anne

Journal of Information Science vol. 24, no. 4, pages 231-240

Publication Date: June 1998

ISSN: 1352-7460

Language: English

Document Type: Journal Article

Record Type: Abstract

Journal Announcement: 3309

Presents a case study of COPAC, a union catalog with access to the online catalog records of some of the largest academic research libraries in the U.K. and Ireland. Points out that COPAC, like all union catalogs, is supplied with multiple copies of records representing the same document in the contributing library catalogs, and that it is necessary to develop duplicate detection and record consolidation procedures in order to reduce the level of duplication visible to the COPAC user. Focuses on measures taken to carry out the duplicate detection and record consolidation procedures, and discusses problem areas encountered. Describes the general structure of these procedures, thus providing a model of the duplicate record handling mechanisms used in COPAC.

Descriptors: Cataloging; Online catalogs; Union catalogs

Classification Codes and Description: 7.03 (Collection Development and Preparation)

Main Heading: Libraries and Information Services

17/5/29 (Item 3 from file: 202)

DIALOG(R) File 202:Info. Sci. & Tech. Abs.

(c) 2004 EBSCO Publishing. All rts. reserv.

3101040

Method of merging large databases in parallel.

Author(s): Hernandez, M A; Stolfo, S.J.

Patent Number(s): US 5497486

Publication Date: Mar 5, 1996

Language: English

Document Type: Patent

Record Type: Abstract

Journal Announcement: 3100

The semantic integration problem for merging multiple databases of very large size, the merge/purge problem, can be solved by multiple runs of the sorted neighborhood method or the clustering method with small windows, followed by the computation of the transitive closure over the results of each run. The sorted neighborhood method works well under this scheme but is computationally expensive due to the sorting phase. An alternative method is introduced, based on data clustering that reduces the complexity to linear time making multiple runs followed by transitive closure feasible and efficient. A method is provided for identifying duplicate records in a database, each record having at least one field and a plurality of keys, including the steps of sorting the records according to a criteria applied to a first key; comparing a number of consecutive sorted records to each other, wherein the number is less than a number of records in said

database and identify a first group of duplicate records ;
storing the identity of the first group; sorting the records according to a
criteria applied to a second key; comparing a number of consecutive sorted
records to each other; and subjecting the union of the first and second
groups to transitive closure.

Descriptors: **Databases** ; Integration; Parallel processing; Semantics
Classification Codes and Description: 5.04 (Advanced Computing, Parallel
Processing); 6.02 (Bibliographic Search Services, **Databases**)
Main Heading: Information Processing and Control; Information Systems and
Applications

?

?ds

Set	Items	Description
S1	57	CO='ARES'
S2	3	S1 AND (DUPLICAT? OR RECORD? OR DATABASE?)
S3	0	S1 AND IC=G06F?

?show files

File 347:JAPIO Nov 1976-2003/Nov(Updated 040308)

(c) 2004 JPO & JAPIO

File 348:EUROPEAN PATENTS 1978-2004/Mar W03

(c) 2004 European Patent Office

File 349:PCT FULLTEXT 1979-2002/UB=20040325,UT=20040318

(c) 2004 WIPO/Univentio

File 350:Derwent WPIX 1963-2004/UD,UM &UP=200419

(c) 2004 Thomson Derwent

?ds

Set	Items	Description
S1	63	AU='ROTTER, J. M.':AU='ROTTER, J.M.'
S2	5	AU='BROWN, BARBARA A.'
S3	26	AU='BROWN, B. C.'
S4	89	S1 OR S3
S5	0	S4 AND (RECORD? OR DUPLICAT?)

?show files

File 35:Dissertation Abs Online 1861-2004/Feb
(c) 2004 ProQuest Info&Learning

File 65:Inside Conferences 1993-2004/Mar W4
(c) 2004 BLDSC all rts. reserv.

File 148:Gale Group Trade & Industry DB 1976-2004/Mar 29
(c)2004 The Gale Group

File 2:INSPEC 1969-2004/Mar W3
(c) 2004 Institution of Electrical Engineers

File 16:Gale Group PROMT(R) 1990-2004/Mar 29
(c) 2004 The Gale Group

File 636:Gale Group Newsletter DB(TM) 1987-2004/Mar 29
(c) 2004 The Gale Group

File 624:McGraw-Hill Publications 1985-2004/Mar 29
(c) 2004 McGraw-Hill Co. Inc

File 275:Gale Group Computer DB(TM) 1983-2004/Mar 29
(c) 2004 The Gale Group

File 647:CMP Computer Fulltext 1988-2004/Mar W3
(c) 2004 CMP Media, LLC

File 674:Computer News Fulltext 1989-2004/Mar W3
(c) 2004 IDG Communications

?

?ds

"

Set	Items	Description
S1	3	AU='ROTTER JOANN MOLARO' OR AU='ROTTER J M'
S2	6	AU='BROWN BARBARA CLAIRE' OR AU='BROWN B C'

SYSTEM:OS - DIALOG OneSearch

.. File 347:JAPIO Nov 1976-2003/Nov(Updated 040308)

(c) 2004 JPO & JAPIO

***File 347: JAPIO data problems with year 2000 records are now fixed.**

Alerts have been run. See HELP NEWS 347 for details.

File 350:Derwent WPIX 1963-2004/UD,UM &UP=200419

(c) 2004 Thomson Derwent

***File 350: For more current information, include File 331 in your search.**

Enter HELP NEWS 331 for details.

File 348:EUROPEAN PATENTS 1978-2004/Mar W03

(c) 2004 European Patent Office

File 349:PCT FULLTEXT 1979-2002/UB=20040325,UT=20040318

(c) 2004 WIPO/Univentio

?t sl/9/1

1/9/1 (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

015441395 **Image available**
WPI Acc No: 2003-503537/200347
XRPX Acc No: N03-399728

Multiple records consolidating method in various organizations and enterprises, involves merging two identified records based on data obtained by comparison of identified records

Patent Assignee: SIEMENS MEDICAL SOLUTIONS HEALTH SERVICE (SIEI)

Inventor: BROWN B C; ROTTER J M

Number of Countries: 026 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20030046280	A1	20030306	US 2001317152	P	20010905	200347 B
			US 2001993041	A	20011106	
WO 200321485	A2	20030313	WO 2002US27501	A	20020828	200347

Priority Applications (No Type Date): US 2001317152 P 20010905; US
2001993041 A 20011106

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20030046280	A1		29	G06F-007/00	Provisional application US 2001317152
WO 200321485	A2	E		G06F-017/30	

Designated States (National): CA JP
Designated States (Regional): AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
IE IT LU MC NL PT SE SK TR

Abstract (Basic): US 20030046280 A1

NOVELTY - The method involves identifying two records. A record matching criteria is then applied to compare the data in the two identified records to determine a common data. The data common in both the records are merged together into a composite record in response to the determined common data.

USE - Used for consolidating multiple records in various organizations and enterprises e.g. health care field.

ADVANTAGE - The redundant and duplicate data are identified effectively and consolidated into a single entity and stored ones repository. This method allows the space to be increased in the memory by deleting all the unwanted data stored. The record about a patient is maintained as unique one if the patient has one or more records, thereby limiting the problem of receiving the wrong treatment.

DESCRIPTION OF DRAWING(S) - The drawing shows a system including an application for preventing the creation of duplicate records and for consolidating multiple records that are associated with a single entity and managing the associated workload.

Record manager application (200)

Duplicate record report process (205)

Web browser (210)

Electronic master person index (213)

Clinical data repository. (217)

pp; 29 DwgNo 2/21

Title Terms: MULTIPLE; RECORD; CONSOLIDATE; METHOD; VARIOUS; MERGE; TWO;
IDENTIFY; RECORD; BASED; DATA; OBTAIN; COMPARE; IDENTIFY; RECORD

Derwent Class: S05; T01

International Patent Class (Main): G06F-007/00; G06F-017/30

File Segment: EPI

Manual Codes (EPI/S-X): S05-G02G; T01-J05B4M; T01-J06A1

?

?t s3/3/1-6

3/3/1 (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

015441395 **Image available**
WPI Acc No: 2003-503537/200347
XRPX Acc No: N03-399728

Multiple records consolidating method in various organizations and enterprises, involves merging two identified records based on data obtained by comparison of identified records

Patent Assignee: SIEMENS MEDICAL SOLUTIONS HEALTH SERVICE (SIEI)

Inventor: BROWN B C ; ROTTER J M

Number of Countries: 026 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20030046280	A1	20030306	US 2001317152	P	20010905	200347 B
			US 2001993041	A	20011106	
WO 200321485	A2	20030313	WO 2002US27501	A	20020828	200347

Priority Applications (No Type Date): US 2001317152 P 20010905; US 2001993041 A 20011106

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20030046280	A1		29	G06F-007/00	Provisional application US 2001317152
WO 200321485	A2	E		G06F-017/30	

Designated States (National): CA JP
Designated States (Regional): AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LU MC NL PT SE SK TR

3/3/2 (Item 2 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

014958384
WPI Acc No: 2003-018898/200301
XRAM Acc No: C03-004665

New isolated insect nuclear receptor polypeptide and encoding nucleic acid, useful for screening assays to identify insecticidal compounds, and can be used as components of a chimeric cassette for inducible gene expression

Patent Assignee: SYNGENTA PARTICIPATIONS AG (SYGN)

Inventor: BROADUS J A; BROWN B C ; KAMDAR K P; STAM L F

Number of Countries: 100 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 200277157	A2	20021003	WO 2002US11257	A	20020322	200301 B

Priority Applications (No Type Date): US 2001278336 P 20010323

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 200277157	A2	E	204	C12N-000/00	

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM ZW
Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZM ZW

3/3/3 (Item 3 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

013294456 **Image available**

WPI Acc No: 2000-466391/200041

XRAM Acc No: C00-140516

XRPX Acc No: N00-348081

Kit for disposing of contaminated intravenous needle and delivering adhesive wound dressings to patient includes adhesive mounting substrate for attaching apparatus to surface

Patent Assignee: BROWN B C (BROW-I); BROWN W S (BROW-I)

Inventor: **BROWN B C** ; BROWN W S

Number of Countries: 002 Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
CA 2284210	A1	20000328	CA 2284210	A	19990928	200041 B
US 6099509	A	20000808	US 98102015	P	19980928	200044
			US 99295486	A	19990420	
US 6302867	B1	20011016	US 98102015	P	19980928	200164
			US 99295486	A	19990420	
			US 2000562601	A	20000501	

Priority Applications (No Type Date): US 99295486 A 19990420; US 98102015 P 19980928; US 2000562601 A 20000501

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
CA 2284210	A1	E	50	A61M-025/02	
US 6099509	A			A61M-005/32	Provisional application US 98102015
US 6302867	B1			A61M-005/32	Provisional application US 98102015
					Cont of application US 99295486
					Cont of patent US 6099509

3/3/4 (Item 4 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

003879375

WPI Acc No: 1984-024913/198405

XRPX Acc No: N84-018744

Wear monitor for steel-reinforced conveyor belt - has transducers spring-loaded towards belt surfaces with surface slide shoes or rollers

Patent Assignee: BROWN B C (BROW-I)

Inventor: **BROWN B C**

Number of Countries: 003 Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
DE 3326232	A	19840126	DE 3326232	A	19830721	198405 B
AU 8316864	A	19840202				198412
ZA 8305296	A	19841116				198513

Priority Applications (No Type Date): AU 825045 A 19820726

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
DE 3326232	A		15		

3/3/5 (Item 1 from file: 348)

DIALOG(R)File 348:EUROPEAN PATENTS

(c) 2004 European Patent Office. All rts. reserv.

01576342

A SYSTEM FOR PROCESSING AND CONSOLIDATING RECORDS

SYSTEM ZUM VERARBEITEN UND KONSOLIDIEREN VON DATENSATZEN

SYSTEMES DE TRAITEMENT ET DE REGROUPEMENT D'ENREGISTREMENTS

PATENT ASSIGNEE:

Siemens Medical Solutions Health Services Corporation, (4092280), 51,
Valley Stream Parkway, Malvern, PA 19355, (US), (Applicant designated
States: all)

INVENTOR:

ROTTER, Joann Molaro , 32 Huntingdon Farm Drive, Glen Mills, PA 19342,
(US)

BROWN, Barbara Claire , 122 Piqua Circle, Malvern, PA 19312, (US)
PATENT (CC, No, Kind, Date):

WO 2003021485 030313

APPLICATION (CC, No, Date): EP 2002757448 020828; WO 2002US27501 020828

PRIORITY (CC, No, Date): US 317152 P 010905; US 993041 011106

DESIGNATED STATES: AT; BE; BG; CH; CY; CZ; DE; DK; EE; ES; FI; FR; GB; GR;
IE; IT; LI; LU; MC; NL; PT

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI

INTERNATIONAL PATENT CLASS: G06F-017/30

LANGUAGE (Publication,Procedural,Application): English; English; English

3/3/6 (Item 1 from file: 349)

DIALOG(R)File 349:PCT FULLTEXT

(c) 2004 WIPO/Univentio. All rts. reserv.

00991449 **Image available**

A SYSTEM FOR PROCESSING AND CONSOLIDATING RECORDS

SYSTEMES DE TRAITEMENT ET DE REGROUPEMENT D'ENREGISTREMENTS

Patent Applicant/Assignee:

SIEMENS MEDICAL SOLUTIONS HEALTH SERVICES CORPORATION, 51 Valley Stream
Parkway, Malvern, PA 19355, US, US (Residence), US (Nationality)

Inventor(s):

ROTTER Joann Molaro , 32 Huntingdon Farm Drive, Glen Mills, PA 19342, US

BROWN Barbara Claire , 122 Piqua Circle, Malvern, PA 19312, US

Legal Representative:

BURKE Alexander J (et al) (agent), Siemens Corporation - Intellectual
Property Dept., 186 Wood Ave. South, Iselin, NJ 08830, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200321485 A2-A3 20030313 (WO 0321485)

Application: WO 2002US27501 20020828 (PCT/WO US02027501)

Priority Application: US 2001317152 20010905; US 2001993041 20011106

Designated States: CA JP

(EP) AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LU MC NL PT SE SK TR

Publication Language: English

Filing Language: English

Fulltext Word Count: 6398

?

?ds

Set	Items	Description
S1	9123135	DUPLICAT? OR DUPE? ? OR COMMON? OR SAME? OR REDUNDAN? OR R- EPLICA? OR REPEAT? OR IDENTICAL? OR DOUBLE? OR ANALOGOUS? OR - SECOND OR PARALLEL? OR MULTIPLE?
S2	6219980	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?.
S3	6237038	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?
S4	15652484	REMOV? OR DELETE? OR DELETING? OR ERASE? ? OR ERASING? OR - DEDUPE? ? OR CHECK? OR FIND???? OR IDENTIF? OR DETECT? OR C- OMPAR???? OR MATCH? OR PREVENT?
S5	112313	S1(3N)S3
S6	8907	S5(5N)S4
S7	7945496	MERGE? OR MERGING? OR CONSOLIDAT? OR COMBIN? OR JOIN??? OR GROUP? OR COMPRESS? OR COMPOSITE?
S8	4330	S7(5N)S5
S9	420	S6 AND S8
S10	19132178	MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL? OR CLINICAL?
S11	218	S10 AND S9
S12	2	S11 AND (EMR OR ELECTRONIC()MEDICAL()RECORD?)
S13	218	S11
S14	119	RD (unique items)
S15	99	S14 NOT PY>2001
S16	26178	15 AND (DATABASE? OR DATA()BASE? OR DATA(3N) (WAREHOUS? OR - WARE()HOUS? OR MINE? ? OR MINING?) OR DATAMIN? OR DB OR DBS OR RECORD?()MANAGE?)
S17	6	S15 AND (DATABASE? OR DATA()BASE? OR DATA(3N) (WAREHOUS? OR WARE()HOUS? OR MINE? ? OR MINING?) OR DATAMIN? OR DB OR DBS OR RECORD?()MANAGE?)

?show files

File 5:Biosis Previews(R) 1969-2004/Mar W3

(c) 2004 BIOSIS

File 73:EMBASE 1974-2004/Mar W3

(c) 2004 Elsevier Science B.V.

File 155:MEDLINE(R) 1966-2004/Mar W4

(c) format only 2004 The Dialog Corp.

File 34:SciSearch(R) Cited Ref Sci 1990-2004/Mar W3

(c) 2004 Inst for Sci Info

File 434:SciSearch(R) Cited Ref Sci 1974-1989/Dec

(c) 1998 Inst for Sci Info

?

12/5/2 (Item 1 from file: 34)

DIALOG(R)File 34:SciSearch(R) Cited Ref Sci
(c) 2004 Inst for Sci Info. All rts. reserv.

09778452 Genuine Article#: 449JL Number of References: 27

Title: LifeCode - A deployed application for automated medical coding

Author(s): Heinze DT; Morsch ML; Sheffer RE; Jimmink MA; Jennings MA;

Morris WC; Morsch AEW

Journal: AI MAGAZINE, 2001, V22, N2 (SUM), P76-88

ISSN: 0738-4602 Publication date: 20010600

Publisher: AMER ASSOC ARTIFICIAL INTELL, 445 BURGESS DRIVE, MENLO PK, CA
94025-3496 USA

Language: English Document Type: ARTICLE

Journal Subject Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE

Abstract: LIFECODE, is a natural language processing (NLP) and expert system that extracts demographic and **clinical** information from free-text **clinical** records. The initial application of LIFECODE. is for the emergency medicine **clinical** specialty. An application for diagnostic radiology went into production in October 2000. The LIFECODE NLP engine uses a large number of specialist readers whose particular output are combined at various levels to form an integrated picture of the **patient** 's **medical** condition(s), course of treatment, and disposition. The LIFECODE expert system performs the tasks of **combining** complementary information, **deleting** redundant information, assessing the level of **medical** risk and level of service represented in the **clinical** record, and producing an output that is appropriate for input to an **electronic medical record** (**EMR**) system or a **hospital** information system. Because of the critical nature of the tasks, LIFECODE has a unique "self-awareness" feature that enables it to recognize the limits of its competence and, thus, ask for assistance from a human expert when faced with information that is beyond the bounds of its competence. The LIFECODE NLP and expert systems reside in various delivery packages, including online transaction processing, a web browser interface, and an automated speech recognition (ASR) interface.

Identifiers--Keyword Plus(R): LANGUAGE

Cited References:

*AMA, 1999, CURR PROC TERM CPT 2

*MED, 1999, PHYS ICD 9 CM INT CL

ARONOW DB, 1995, IR67 U MASS AMH CTR

ARONOW DB, 1996, P AM MED INF ASS FAL

ARONOW DB, 1995, IR61 U MASS AMH CTR

ARONOW DB, 1999, V6, P393, J AM MED INFORM ASSN

BENTLEY J, 1996, V14, P75, UNIX REV

CHRISTENSEN CM, 2000, V78, P102, HARVARD BUS REV

CROFT WB, 1995, P1719, P MEDINFO 95 AMST IO

FREUND JE, 1980, MATH STAT

FRIEDMAN N, 1997, V29, P131, MACH LEARN

HIRSCH M, 1995, IR63 U MASS AMH CTR

LANGACKER RW, 1999, C NAT EXPL LING THEO

LARKEY LS, 1995, IR64 U MASS AMH CTR

LEHNERT W, 1994, TC32 U MASS AMH CTR

LENERT LA, 1993, P274, P ANN S COMP APPL ME

MORRIS WC, 2000, CD ROM P AM MED INF

RANUM DL, 1988, P141, P 12 ANN S COMP APPL

SAGER N, 1994, P230, P S COMP APPL MED CA

SAGER N, 1994, V1, P142, J AM MED INFORM ASSN

SAGER N, 1996, P547, P AMIA ANN FALL S 96

SNEIDERMAN C, 1996, P239, P AMIA ANN FALL S

SNEIDERMAN CA, 1995, CD ROM P AM MED INF

SONDERLAND S, 1995, TC39 U MASS AMH CTR

WICKENS CD, 1992, ENG PSYCHOL HUMAN PE

YANG Y, 1992, P460, P 16 ANN S COMP APPL

ZINGMOND D, 1993, V26, P467, COMPUT BIOMED RES

?ds

Set	Items	Description
S1	4642259	DUPLICAT? OR DUPE? ? OR COMMON? OR SAME? OR REDUNDAN? OR R- EPLICA? OR REPEAT? OR IDENTICAL? OR DOUBLE? OR ANALOGOUS? OR - SECOND OR PARALLEL? OR MULTIPLE?
S2	2897100	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?.
S3	2897155	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?
S4	6310996	REMOV? OR DELETE? OR DELETING? OR ERASE? ? OR ERASING? OR - DEDUPE? ? OR CHECK? OR FIND???? OR IDENTIF? OR DETECT? OR C- OMPAR???? OR MATCH? OR PREVENT?
S5	158199	S1(3N)S3
S6	14866	S5(5N)S4
S7	2884602	MERGE? OR MERGING? OR CONSOLIDAT? OR COMBIN? OR JOIN??? OR GROUP? OR COMPRESS? OR COMPOSITE?
S8	6327	S7(5N)S5
S9	643	S6 AND S8
S10	455519	MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL? OR CLINICAL?
S11	16	S10 AND S9
S12	48	S9 AND IC=(G06F-007/00 OR G06F-017/30)
S13	7	S12 AND MC=(S05-G02G OR T01-J05B4M OR T01-J06A1)
S14	39	S12 NOT (S13 OR S11)

?show files

File 347:JAPIO Nov 1976-2003/Nov(Updated 040308)

(c) 2004 JPO & JAPIO

File 350:Derwent WPIX 1963-2004/UD,UM &UP=200419

(c) 2004 Thomson Derwent

?

?t' s11/5/9,14

11/5/9 (Item 4 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

015441395 **Image available**
WPI Acc No: 2003-503537/200347
XRPX Acc No: N03-399728

Multiple records consolidating method in various organizations and enterprises, involves merging two identified records based on data obtained by comparison of identified records

Patent Assignee: SIEMENS MEDICAL SOLUTIONS HEALTH SERVICE (SIEI)
Inventor: BROWN B C; ROTTER J M
Number of Countries: 026 Number of Patents: 002
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20030046280	A1	20030306	US 2001317152	P	20010905	200347 B
			US 2001993041	A	20011106	
WO 200321485	A2	20030313	WO 2002US27501	A	20020828	200347

Priority Applications (No Type Date): US 2001317152 P 20010905; US 2001993041 A 20011106

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20030046280	A1		29	G06F-007/00	Provisional application US 2001317152
WO 200321485	A2	E		G06F-017/30	

Designated States (National): CA JP
Designated States (Regional): AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LU MC NL PT SE SK TR

Abstract (Basic): US 20030046280 A1

NOVELTY - The method involves identifying two records. A record matching criteria is then applied to compare the data in the two **identified records** to determine a **common data**. The **data common** in both the **records** are **merged** together into a **composite** record in response to the determined common data.

USE - Used for **consolidating multiple records** in various organizations and enterprises e.g. **health care field**.

ADVANTAGE - The **redundant and duplicate data** are **identified** effectively and **consolidated** into a single entity and stored ones repository. This method allows the space to be increased in the memory by deleting all the unwanted data stored. The record about a **patient** is maintained as unique one if the **patient** has one or more records, thereby limiting the problem of receiving the wrong treatment.

DESCRIPTION OF DRAWING(S) - The drawing shows a system including an application for **preventing** the creation of **duplicate records** and for **consolidating multiple records** that are associated with a single entity and managing the associated workload.

Record manager application (200)
Duplicate record report process (205)
Web browser (210)
Electronic master person index (213)
Clinical data repository. (217)
pp; 29 DwgNo 2/21

Title Terms: MULTIPLE; RECORD; CONSOLIDATE; METHOD; VARIOUS; MERGE; TWO; IDENTIFY; RECORD; BASED; DATA; OBTAIN; COMPARE; IDENTIFY; RECORD
Derwent Class: S05; T01
International Patent Class (Main): G06F-007/00; G06F-017/30
File Segment: EPI

11/5/14 (Item 9 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

013815283 **Image available**

WPT Acc No: 2001-299495/131
XRPX Acc No: N01-214804

Electronic medical record registry system displays registered electronic medical record from main registry database to collect and store information of patient histories from service provider databases
Patent Assignee: UNIV SAN DIEGO STATE FOUND (UYSA-N); SOFTWARE PARTNERS (SOFT-N)

Inventor: KIRNAK A

Number of Countries: 092 Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 200065522	A2	20001102	WO 2000US11390	A	20000428	200131 B
AU 200048054	A	20001110	AU 200048054	A	20000428	200131
EP 1145180	A2	20011017	EP 2000930190	A	20000428	200169
			WO 2000US11390	A	20000428	

Priority Applications (No Type Date): US 99131434 P 19990428

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
-----------	------	--------	----------	--------------

WO 200065522	A2	E	28	G06F-019/00
--------------	----	---	----	-------------

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ TZ UG ZW

AU 200048054	A		G06F-019/00	Based on patent WO 200065522
--------------	---	--	-------------	------------------------------

EP 1145180	A2	E	G06F-019/00	Based on patent WO 200065522
------------	----	---	-------------	------------------------------

Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI

Abstract (Basic): WO 200065522 A2

NOVELTY - The system has display programs (204a,204b) to display registered **medical** records. The **medical** service provider databases (200) store electronic **medical** records having information indicating histories of **patients** associated with the respective service provider. A registry repository includes a main registry database (104) with registered electronic **medical** records.

DETAILED DESCRIPTION - The system includes a **match / merge** program (206) which **duplicates** the electronic **medical records** associated with the **medical** service provider databases (200), based on unique **patient** information, to form the registered **medical** record of complete **patient** history. The program assigns a unique **patient** identifier to each **patient** and assigns the same identifier to each **medical** service provider database record stored in the main registry database (104). A standard data exchange format is used between the main registry database and the service provider databases to transfer **medical** records in a standardized form. An INDEPENDENT CLAIM is also included for method of storing and accessing electronic **medical** records.

USE - For transferring electronic **medical** records from heterogeneous databases through computers across local networks and internet, into a central registry database to merge the records for presentation.

ADVANTAGE - Allows all service care providers to become part of a network that is updated with immunization information from both private and public databases which are willing to share electronic **medical** records. The electronic records are processed through a **match / merge** algorithm, to identify the **multiple** sourced **records** belonging to a single **patient**, and provide a completed **patient** history at the point of service.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram illustrating the basic architecture for the flow of information and the interfaces to the main registry database.

Databases (104,200)

Display programs (204a,204b)

Match/merge program (206)

pp; 28 DwgNo 2/1

Title Terms: ELECTRONIC; **MEDICAL** ; RECORD; REGISTER; SYSTEM; DISPLAY;
REGISTER; ELECTRONIC; **MEDICAL** ; RECORD; MAIN; REGISTER; DATABASE;
COLLECT; STORAGE; INFORMATION; **PATIENT** ; SERVICE

Derwent Class: T01

International Patent Class (Main): G06F-019/00

File Segment: EPI

?

13/5/1 (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

015770436 **Image available**
WPI Acc No: 2003-832638/200377
XRPX Acc No: N03-665593

Record classification method, for records for delivery to different destinations, involves forming record group and assigning destination identification to each replica record except one record in that group

Patent Assignee: IBM CANADA LTD (IBMC); COCHRANE R J (COCH-I); LAPIS G (LAPI-I); PIRAHESH H (PIRA-I); RIELAU S (RIEL-I); SIDLE R S (SIDL-I); ZHU B (ZHUB-I)

Inventor: COCHRANE R J; LAPIS G; PIRAHESH H; RIELAU S P; SIDLE R S; ZHU B; RIELAU S

Number of Countries: 002 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20030204540	A1	20031030	US 2003393852	A	20030321	200377 B
CA 2384174	A1	20031029	CA 2384174	A	20020429	200378

Priority Applications (No Type Date): CA 2384174 A 20020429

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20030204540	A1		19	G06F-012/00	
CA 2384174	A1	E		G06F-017/30	

Abstract (Basic): US 20030204540 A1

NOVELTY - A **group of replica records** is formed for a record among selected subset of records. A destination **identification** is assigned for each **replica record** in the **group** except one **replica record**. A **replica record** associated with one of several destinations, is selected and placed in that destination.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) computer readable medium storing instructions for selecting subset of records;

(2) data processing system.

USE - For classifying records for delivery to different targets or destinations.

ADVANTAGE - Records of information are classified for delivery to different targets or destinations reliably, even though it is not predictable as to which target, a given unit of data is to be routed.

DESCRIPTION OF DRAWING(S) - The figure shows a block diagram of the record classification system.

records (102)

database management system (106)

computer system (116)

memory (118)

network computer (128)

pp; 19 DwgNo 1/4

Title Terms: RECORD; CLASSIFY; METHOD; RECORD; DELIVER; DESTINATION; FORMING; RECORD; GROUP; ASSIGN; DESTINATION; IDENTIFY; REPLICA; RECORD; ONE; RECORD; GROUP

Derwent Class: T01

International Patent Class (Main): G06F-012/00; G06F-017/30

International Patent Class (Additional): G06F-012/02

File Segment: EPI

13/5/6 (Item 6 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

012850167 **Image available**

WPI Acc No: 2000-021999/200002

Related WPI Acc No: 1998-322133

XRPX Acc No: N00-016273

Dynamic and hybrid sparse indices generating method for workfiles used in SQL queries in relational database management system

Patent Assignee: INT BUSINESS MACHINES CORP (IBMC)

Inventor: BHARGAVA G; DESAI P S; GOEL P; HOA P; IYER B R; LIN F; MUKAI J; PERLMAN W S; TIE H S

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 5978792	A	19991102	US 95393803	A	19950224	200002 B
			US 986745	A	19980113	

Priority Applications (No Type Date): US 95393803 A 19950224; US 986745 A 19980113

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 5978792	A		25	G06F-017/30	Cont of application US 95393803 Cont of patent US 5758145

Abstract (Basic): US 5978792 A

NOVELTY - A sparse index stores entries containing key value that indicates approximate location of data records. Sparse index is probed to determine starting location of next search of key value in the query. An entry is returned from the sparse index without performing next search, when it matches with the key value, otherwise next search is performed.

DETAILED DESCRIPTION - The sparse index comprises a fixed number of key values along with their record **identifiers** (RIDs) (410) in a **file** . **Duplicate** entries are eliminated during the **merge** phase of file creation. A dynamic portion (404) is created for sparse index during the search phase of file (406) access. The exact location of the data record in the file, is determined using an auxiliary search and the approximate location provided by the sparse index. An INDEPENDENT CLAIM is also included for query performing apparatus.

USE - For RDBMS using dynamic and hybrid sparse index, in mainframe, minicomputer and personal computers.

ADVANTAGE - The performance of workfiles used in SQL queries is optimized by searching data record using sparse index.

DESCRIPTION OF DRAWING(S) - The figure depicts the block diagram illustrating hybrid sparse index comprising static and dynamic portion.

Dynamic portion (404)

File (406)

Record identifiers (410)

pp; 25 DwgNo 4/6

Title Terms: DYNAMIC; HYBRID; INDEX; GENERATE; METHOD; SQL; QUERY; RELATED; DATABASE; MANAGEMENT; SYSTEM

Derwent Class: T01

International Patent Class (Main): G06F-017/30

File Segment: EPI

13/5/7 (Item 7 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

011674964 **Image available**

WPI Acc No: 1998-091873/199809

XRPX Acc No: N98-073106

Database updating method for client and server type database system interconnected via network - by performing combining process of updating record for record registered into database and which has same identification

Patent Assignee: MATSUSHITA GRAPHIC COMMUNICATION SYSTEMS (MATY)

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 9319635	A	19971212	JP 96136233	A	19960530	199809 B

Priority Applications (No Type Date): JP 96136233 A 19960530

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
JP 9319635	A		15	G06F-012/00	

Abstract (Basic): JP 9319635 A

The method involves sequentially receiving updating indication and updating record for a record registered into a database. A combining process is performed to the updating **record** which has the **same identification**. After the **combining** process, the combined updating record is used to update the database record.

ADVANTAGE - Facilitates batch processing and updating process of database record. Effectively utilises memory which stores updating record, by erasing unnecessary old record at every combining process execution. Reduces load of batch processing by reducing number of batch records. Enables optimum implementation of whole client and server type database system.

Dwg.1/17

Title Terms: DATABASE; UPDATE; METHOD; CLIENT; SERVE; TYPE; DATABASE; SYSTEM; INTERCONNECT; NETWORK; PERFORMANCE; COMBINATION; PROCESS; UPDATE; RECORD; RECORD; REGISTER; DATABASE; IDENTIFY

Derwent Class: T01

International Patent Class (Main): G06F-012/00

International Patent Class (Additional): **G06F-017/30**

File Segment: EPI

?

t s14/5/4,5,8,9,13,14,16,23,28,29,31,32,39

14/5/4 (Item 4 from file: 347)
DIALOG(R) File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

07483379 **Image available**
PROGRAM FOR PREDICTING INFORMATION USE FREQUENCY, DEVICE FOR PREDICTING
INFORMATION USE FREQUENCY AND METHOD FOR PREDICTING INFORMATION USE
FREQUENCY

PUB. NO.: 2002-351897 [JP 2002351897 A]
PUBLISHED: December 06, 2002 (20021206)
INVENTOR(s): UKAI TAKANORI
MIMATSU KAZUO
APPLICANT(s): FUJITSU LTD
APPL. NO.: 2001-153116 [JP 2001153116]
FILED: May 22, 2001 (20010522)
INTL CLASS: G06F-017/30 ; G06F-017/60

ABSTRACT

PROBLEM TO BE SOLVED: To predict information use frequency at high accuracy.

SOLUTION: In a program, a **comparison** part where a **second** time-series **data** D20 concerning the frequency of use of 'new year postcards' as a separate keyword with a first time-series data D10 concerning the frequency of use of 'camera' as the keyword inputted in a keyword search of the Internet as a reference is sequentially shifted in units of specified time on a time axis and the relative coefficient of the first time-series data D10 and the second time-series data D20 for each specified time unit is calculated and a prediction part specifying a **group** of the first time-series **data** D10 and the **second** time-series **data** D20 corresponding to the relative coefficient with the greatest value among the multiple number of relative coefficients calculated by the comparison part and predicting the frequency of use of the first time-series data D10 constituting the **group** based on the **second** time-series **data** D20 are provided.

COPYRIGHT: (C)2003,JPO

14/5/5 (Item 5 from file: 347)
DIALOG(R) File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

06795294 **Image available**
INFORMATION RETRIEVAL DEVICE, INFORMATION COMPRESSING METHOD FOR
INFORMATION RETRIEVAL DEVICE, AND RECORDING MEDIUM

PUB. NO.: 2001-022775 [JP 2001022775 A]
PUBLISHED: January 26, 2001 (20010126)
INVENTOR(s): ITO MASAO
APPLICANT(s): MATSUSHITA ELECTRIC IND CO LTD
APPL. NO.: 11-194740 [JP 99194740]
FILED: July 08, 1999 (19990708)
INTL CLASS: G06F-017/30 ; G06F-005/00; H03M-007/30

ABSTRACT

PROBLEM TO BE SOLVED: To suppress the capacity of an index file by compressing index information more efficiently and to enable high-speed retrieval using the index file.

SOLUTION: When entries of index **information** are successively **identical**, this **information** **compressing** method **deletes** by an index information compression part 107 the entries except only one entry, puts (n) successive pieces of index **information** with the **same** document number in one **group**

as to index **information** in the **same** character **ch**, and generates **compressed** index information having the document number compressed into one. When byte data of position numbers are not larger than a reference value (0×0f), the low-order four bits of each of the byte data of the position numbers are connected to generate compressed index information having the position number converted into 1-byte data.

COPYRIGHT: (C)2001,JPO

14/5/8 (Item 8 from file: 347)
DIALOG(R)File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

05421947 **Image available**
DATA COMPRESSION METHOD AND DATA COMPRESSOR

PUB. NO.: 09-036747 [JP 9036747 A]
PUBLISHED: February 07, 1997 (19970207)
INVENTOR(s): SAKAKIBARA HIDETOSHI
TODA ATSUKO
APPLICANT(s): TOSHIBA CORP [000307] (A Japanese Company or Corporation), JP
(Japan)
TOSHIBA COMPUT ENG CORP [486760] (A Japanese Company or
Corporation), JP (Japan)
APPL. NO.: 07-181485 [JP 95181485]
FILED: July 18, 1995 (19950718)
INTL CLASS: [6] H03M-007/40; G06F-005/00; **G06F-017/30**
JAPIO CLASS: 42.4 (ELECTRONICS -- Basic Circuits); 45.1 (INFORMATION
PROCESSING -- Arithmetic Sequence Units); 45.4 (INFORMATION
PROCESSING -- Computer Applications)

ABSTRACT

PROBLEM TO BE SOLVED: To provide the data compression method and the data compressor by which data can be compressed at a high speed without taking much time for data retrieval by adopting the hash method for data compression by the LZ slide dictionary method so as to retrieve data.

SOLUTION: When a data input output device 1 reads compression processing object data to a memory 3, a control circuit 2 retrieves the **data** string the **same** as **compression** processing object data string from a hash table by the hash method and compares the data of the retrieved data string with that of the compression processing object data string one by one and provides an output of equal data length of both the **data** strings, **repeats** the retrieval and **comparison** till the retrieval object is not in existence and selects a data string whose equal data length is longest among data strings at the retrieved hand position and succeeding positions and provides an output of the head position of the selected data string and the equal data length as equality information and registers the head position of the compression processing object data string to a hash table as a retrieval object.

14/5/9 (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

015887469 **Image available**
WPI Acc No: 2004-045304/200405
XRPX Acc No: N04-036920

Multimedia data fragmenting method for Internet, involves identifying pair of nodes which minimizes parameter proportional to data common to associated data, for forming definitive cells
Patent Assignee: THOMSON LICENSING SA (CSFC); ROBERT P (ROBE-I); STAUDER J (STAU-I)
Inventor: ROBERT P; STAUDER J
Number of Countries: 032 Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
EP 1367831	A2	20031203	EP 2003101367	A	20030515	200405 B
US 20030225791	A1	20031204	US 2003448235	A	20030529	200405
FR 2840424	A1	20031205	FR 20026878	A	20020530	200405

Priority Applications (No Type Date): FR 20026878 A 20020530

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
EP 1367831	A2	E	9	H04N-007/24	
Designated States (Regional): AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LT LU LV MC MK NL PT RO SE SI SK TR					
US 20030225791	A1			G06F-007/00	
FR 2840424	A1			G06F-017/30	

Abstract (Basic): EP 1367831 A2

NOVELTY - The multimedia data (D'6-D'9) associated with nodes (N'6-N'9) are grouped into a temporary cell. When the temporary cell reaches a threshold size, a pair of nodes (N'7,N'8) which minimizes a parameter proportional to number of **data common** to associated **data** (D'7,D'8) is **identified**, for forming a definitive cells (C'5-C'7) comprising data (D'5-D'7).

DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for multimedia data fragmenting device.

USE - For fragmenting multimedia data associated with navigation nodes in telecommunication network such as Internet.

ADVANTAGE - Optimizes use of memory capacity and computational capacity, by fragmenting multimedia **data** and **grouping identical or redundant data** into **same** cell.

DESCRIPTION OF DRAWING(S) - The figure shows a schematic view of cell comprising nodes associated with multimedia data.

definitive cells (C'5-C'7)

multimedia data (D'5-D'9)

nodes (N'6-N'9)

pp; 9 DwgNo 2/2

Title Terms: DATA; FRAGMENT; METHOD; IDENTIFY; PAIR; NODE; MINIMISE;

PARAMETER; PROPORTION; DATA; COMMON; ASSOCIATE; DATA; FORMING; CELL

Derwent Class: T01

International Patent Class (Main): G06F-007/00 ; G06F-017/30 ;

H04N-007/24

File Segment: EPI

14/5/13 (Item 5 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

014604513 **Image available**

WPI Acc No: 2002-425217/200245

Related WPI Acc No: 2002-626900

XRPX Acc No: N02-334351

Data replication **system for employee management system**, matches **data items of several master tables and replicates** matched data items **to generate replica table using particular data in master table**

Patent Assignee: HITACHI LTD (HITA); YOKOUCHI H (YOKO-I)

Inventor: YOKOUCHI H

Number of Countries: 002 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20020038315	A1	20020328	US 2001809257	A	20010316	200245 B
JP 2002108681	A	20020412	JP 2000294551	A	20000927	200245

Priority Applications (No Type Date): JP 2000294551 A 20000927

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20020038315	A1		14	G06F-017/30	
JP 2002108681	A		12	G06F-012/00	

Abstract (Basic): US 200 38315 A1

NOVELTY - A matching unit matches data items of a number of master tables (108,109) using particular data in the master tables as a joining key to **join** each other. A replicating unit **replicates** the **matched data** items based on the **joining** key to generate one replica table (121).

DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for data replication program.

USE - For replicating data of several master tables of database in employee management system.

ADVANTAGE - The data of number of master tables is replicated in one replica table, and hence job execution time is minimized.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of data replica system.

Master tables (108,109)

Replica table (121)

pp; 14 DwgNo 1/7

Title Terms: DATA; REPLICATION; SYSTEM; EMPLOY; MANAGEMENT; SYSTEM; MATCH; DATA; ITEM; MASTER; TABLE; REPLICATION; MATCH; DATA; ITEM; GENERATE; REPLICATION; TABLE; DATA; MASTER; TABLE

Derwent Class: T01

International Patent Class (Main): G06F-012/00; G06F-017/30

File Segment: EPI

14/5/14 (Item 6 from file: 350)

DIALOG(R) File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

014523676 **Image available**

WPI Acc No: 2002-344379/200238

XRPX Acc No: N02-270998

Distributed database system for distributed system, has addition table and deletion table which are compressed into compression file, using which records of duplication database are added/ deleted synchronously

Patent Assignee: HITACHI SOFTWARE ENG CO LTD (HISF)

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 2002007441	A	20020111	JP 2000191459	A	20000626	200238 B

Priority Applications (No Type Date): JP 2000191459 A 20000626

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes

JP 2002007441 A 10 G06F-017/30

Abstract (Basic): JP 2002007441 A

NOVELTY - An addition table (35) and a deletion table (34) containing updating information of duplication database (11,12) and deletion information of the database respectively, are compressed into a compression file (36). The records in each database corresponding to the tables, are added or deleted synchronously using the compression file.

USE - Distributed database system for use in distributed system.

ADVANTAGE - The records are added/deleted synchronously within short time, thus high speed synchronization process is achieved.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of distributed database system. (Drawing includes non-English language text).

Duplication database (11,12)

Deletion table (34)

Addition table (35)

Compression file (36)

pp; 10 DwgNo 1/7

Title Terms: DISTRIBUTE; DATABASE; SYSTEM; DISTRIBUTE; SYSTEM; ADD; TABLE; DELETE; TABLE; COMPRESS; COMPRESS; FILE; RECORD; DUPLICATE; DATABASE; ADD

r; DELETE; SYNCHRONOUS
Derwent Class: T01
International Patent Class (Main): G06F-017/30
International Patent Class (Additional): G06F-012/00
File Segment: EPI

14/5/16 (Item 8 from file: 350)
DIALOG(R) File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

013956761 **Image available**
WPI Acc No: 2001-440975/200147
Related WPI Acc No: 2000-316941; 2001-406751; 2001-578447; 2001-579868;
2001-625043; 2003-416629; 2003-851941
XRPX Acc No: N01-326212

Method for merging multiple knowledge bases into one optimized and
compressed knowledge base to allow for more efficient use
Patent Assignee: KHEDKAR S S (KHED-I); MITTAL S (MITT-I); SELECTICA INC
(SELE-N)

Inventor: KHEDKAR S S; MITTAL S
Number of Countries: 001 Number of Patents: 002
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20010007989	A1	20010712	US 97962594	A	19971031	200147 B
			US 9837398	A	19980309	
			US 2001770938	A	20010125	
US 6480859	B2	20021112	US 97962594	A	19971031	200278
			US 9837398	A	19980309	
			US 2001770938	A	20010125	

Priority Applications (No Type Date): US 9837398 A 19980309; US 97962594 A
19971031; US 2001770938 A 20010125

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20010007989	A1	12	G06F-017/00		CIP of application US 97962594 Div ex application US 9837398 CIP of patent US 6049822 Div ex patent US 6205446
US 6480859	B2		G06F-017/30		CIP of application US 97962594 Div ex application US 9837398 CIP of patent US 6049822 Div ex patent US 6205446

Abstract (Basic): US 20010007989 A1

NOVELTY - The system works by integrating multiple knowledge base files into single files, keeping all of the parameters and constraints of these **multiple files** by identifying two or more **common** parameters among the **files**. These parameters are then **merged** into a single statement in the integrated file, the single statement listing all of the values of the several statements in the multiple files. Additional constraints are added to the new, smaller file to properly relate the values to a characteristic of the original files. Combining the value and activity constraints into a single constraint statement optimizes the resulting single file.

DETAILED DESCRIPTION - An INDEPENDENT CLAIM is included for a software engine for integrating a first and a second knowledge base file into a single smaller integrated third file.

USE - For generating, merging, and optimizing Knowledge Bases.

ADVANTAGE - The method allows the user to go through the selection and configuration process using the bottom-up approach, speeding the process up.

DESCRIPTION OF DRAWING(S) - The figure shown is a flow chart illustrating a knowledge base merging and optimization process.

pp; 12 DwgNo 2/4

Title Terms: METHOD; MERGE; MULTIPLE; BASE; ONE; OPTIMUM; COMPRESS; BASE;
ALLOW; MORE; EFFICIENCY

Derwent Class: T01
International Patent Class (Main): G06F-017/00; G06F-017/30
File Segment: EPI

14/5/21 (Item 13 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

013162934 **Image available**
WPI Acc No: 2000-334807/200029
XRPX Acc No: N00-252498

**Document display method for newspaper report, received via Internet,
involves comparing similarity between chosen document and original report
and displaying combined common data and original information**

Patent Assignee: HITACHI LTD (HITA)
Number of Countries: 001 Number of Patents: 001
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 2000105769	A	20000411	JP 98273494	A	1998092	200029 B

Priority Applications (No Type Date): JP 98273494 A 19980928

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
JP 2000105769	A		26	G06F-017/30	

Abstract (Basic): JP 2000105769 A

NOVELTY - The similarity between multiple reports is evaluated and then documents are divided into multiple portions. The similarity between each divided portion is evaluated and one document is chosen. The information common to chosen document and original report is **detected**. The **common information** and original **information** are **combined** and displayed.

USE - For newspaper report, magazine report received through Internet.

ADVANTAGE - Enables detection and display of original document, thereby enabling efficient perusal of report by user.

pp; 26 DwgNo 1/24

Title Terms: DOCUMENT; DISPLAY; METHOD; NEWSPAPER; REPORT; RECEIVE; COMPARE
; SIMILAR; CHOICE; DOCUMENT; ORIGINAL; REPORT; DISPLAY; COMBINATION;
COMMON; DATA; ORIGINAL; INFORMATION

Derwent Class: T01
International Patent Class (Main): G06F-017/30
File Segment: EPI

14/5/23 (Item 15 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

012637680 **Image available**
WPI Acc No: 1999-443784/199937
XRPX Acc No: N99-331004

**Multilevel source data structure compression method used in field of
document management in general purpose computer system**

Patent Assignee: MICROSOFT CORP (MICR-N)
Inventor: ROSS B E
Number of Countries: 001 Number of Patents: 001
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 5933842	A	19990803	US 96660149	A	19960523	199937 B

Priority Applications (No Type Date): US 96660149 A 19960523

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 5933842	A		20	G06F-017/30	

Abstract (Basic): US 593 02 A

NOVELTY - The reference in the constituent data structure at the next higher level is replaced with a reference that refers to a constituent data structure stored in a destination data structure. The matching constituent data structures are identified. The links to the deleted constituent data structures are then replaced with links to the selected constituent data structure.

DETAILED DESCRIPTION - The data structure at the next higher level refers to a constituent data structure at the selected level of the source data structure. Each constituent data structure has specific contents. All of the constituent **data** structures having the **same** contents are **identified** as **matching**.

USE - For compressing multilevel source data structure for document management in general purpose computer system.

ADVANTAGE - Since a reference is inserted in the compressed document set, that refers to the matching constituent data structure already contained in the **compressed** document set, avoids existence of **duplicate** constituent **data** structure in the **compressed** document set.

DESCRIPTION OF DRAWING(S) - The figure shows the flow diagram of compress documents routine.

pp; 20 DwgNo 6A/8

Title Terms: MULTILEVEL; SOURCE; DATA; STRUCTURE; COMPRESS; METHOD; FIELD; DOCUMENT; MANAGEMENT; GENERAL; PURPOSE; COMPUTER; SYSTEM

Derwent Class: T01

International Patent Class (Main): G06F-017/30

File Segment: EPI

14/5/28 (Item 20 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

012001830 **Image available**

WPI Acc No: 1998-418740/199836

XRPX Acc No: N98-326470

Data processing system for printing of goods catalogue - searches grouped object data corresponding to input search data and outputs searched object data by output device

Patent Assignee: TOPPAN PRINTING CO LTD (TOPP)

Number of Countries: 001 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 10171816	A	19980626	JP 96342704	A	19961206	199836 B
JP 3424473	B2	20030707	JP 96342704	A	19961206	200345

Priority Applications (No Type Date): JP 96342704 A 19961206

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
-----------	------	--------	----------	--------------

JP 10171816	A	7	G06F-017/30	
-------------	---	---	-------------	--

JP 3424473	B2	7	G06F-017/30	Previous Publ. patent JP 10171816
------------	----	---	-------------	-----------------------------------

Abstract (Basic): JP 10171816 A

The system stores several object data. These object **data** are divided into **multiple groups**, based on a preset rule and division information of each object data. The object data with which division is performed, is **compared** with an **identification data**. If they are **identical** then the **matched data** are stored.

Input search data corresponding to attribute of each object data that are matching with identification data, are stored separately. Then, object data corresponding to input search data belonging to respecting group are searched. These searched data are then output, by an output device (3).

ADVANTAGE - Reads out object data, efficiently.

Dwg.1/8

Title Terms: DATA; PROCESS; SYSTEM; PRINT; GOODS; CATALOGUE; SEARCH; GROUP; OBJECT; DATA; CORRESPOND; INPUT; SEARCH; DATA; OUTPUT; SEARCH; OBJECT;

DATA; OUTPUT; DEVICE
Derwent Class: T01
International Patent Class (Main): G06F-017/30
File Segment: EPI

14/5/29 (Item 21 from file: 350)
DIALOG(R) File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

011864986 **Image available**
WPI Acc No: 1998-281896/199825
XRPX Acc No: N98-222431

Parallel database processing system - forwards intermediate accumulation result from accumulation processing server to enquiry processing server, which unifies various results

Patent Assignee: HITACHI LTD (HITA)
Number of Countries: 001 Number of Patents: 001
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 10097544	A	19980414	JP 96249589	A	19960920	199825 B

Priority Applications (No Type Date): JP 96249589 A 19960920

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
JP 10097544	A	27	G06F-017/30	

Abstract (Basic): JP 10097544 A

The system includes an enquiry process server, an I/O server and an accumulation server connected through a network. The record read by the I/O server is allocated to accumulation process server by a record allocation unit. The fields of a record are divided into groups and a group identifier is assigned to each **group**. Several **records** with **same group identifier** values are further **grouped** together.

The accumulation of the records is based on the group identifier value. The storing area of the accumulated records is decided, also based on group identifier value. The accumulation result is periodically updated. Intermediate accumulation results are notified to the enquiry process server. The enquiry process server unifies the various intermediate accumulation results.

ADVANTAGE - Realizes efficient database processing.

Dwg.1/23

Title Terms: PARALLEL; DATABASE; PROCESS; SYSTEM; FORWARD; INTERMEDIATE; ACCUMULATE; RESULT; ACCUMULATE; PROCESS; SERVE; ENQUIRY; PROCESS; SERVE; UNIFIED; VARIOUS; RESULT

Derwent Class: T01

International Patent Class (Main): G06F-017/30
International Patent Class (Additional): G06F-007/24
File Segment: EPI

14/5/31 (Item 23 from file: 350)
DIALOG(R) File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

011446639 **Image available**
WPI Acc No: 1997-424546/199739
XRPX Acc No: N97-353703

Data compressing method for data processing system - involves lexically sorting group of N cells so that lexical order of N cells is established based upon lexical value of substring in each of N cells

Patent Assignee: INT BUSINESS MACHINES CORP (IBMC); IBM CORP (IBMC)
Inventor: STROHACKER O C
Number of Countries: 002 Number of Patents: 003
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 5659755	A	19970819	US 95559673	A	19951120	199739 B

JP 9223972 A 19970 JP 96300161 A 19961112 99744
JP 3213244 B2 20011002 JP 96300161 A 19961112 200164

Priority Applications (No Type Date): US 95559673 A 19951120

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 5659755	A		16	G06F-017/30	
JP 9223972	A		17	H03M-007/40	
JP 3213244	B2		17	H03M-007/40	Previous Publ. patent JP 9223972

Abstract (Basic): US 5659755 A

The method involves selecting a block of N bytes in the data processing system, wherein N is greater than 1. A group of N data cells are created, each of the cells has a physical tag, P, for indicating a physical order in the group. The group of N cells are lexically sorted so that a lexical order of the N cells is established based upon the lexical value of the substring in each of the N cells. Matching strings are identified in the substrings of selected ones of the N cells in the lexically sorted group of N cells. The block of N bytes are encoded utilising the identified matching strings to produce a compressed data set. The compressed data set is efficiently produced by **identifying redundant information** in the lexically sorted **group** of N cells.

ADVANTAGE - Quickly and efficiently compress block of data while maximising amount data processing that may be performed in parallel.

Dwg.5/8

Title Terms: DATA; COMPRESS; METHOD; DATA; PROCESS; SYSTEM; SORT; GROUP; N; CELL; SO; LEXICAL; ORDER; N; CELL; ESTABLISH; BASED; LEXICAL; VALUE; N; CELL

Derwent Class: T01; U21

International Patent Class (Main): G06F-017/30 ; H03M-007/40

International Patent Class (Additional): G06F-012/00; H03M-007/30

File Segment: EPI

14/5/32 (Item 24 from file: 350)

DIALOG(R) File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

011293792 **Image available**

WPI Acc No: 1997-271697/199724

XRPX Acc No: N97-225224

Identifying textual documents and multimedia files corresponding to search topic - accepting query and returning single search results list having text and multimedia information

Patent Assignee: INFONAUTICS CORP (INFO-N)

Inventor: BARR T; BEATTIE J T; HUSICK L A; KOPELMAN J; KRUPIT M S; MORGAN H ; WATKEYS E H; WEINBERGER M I

Number of Countries: 024 Number of Patents: 006

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 9710537	A2	19970320	WO 96US15664	A	19960913	199724 B
AU 9672026	A	19970401	AU 9672026	A	19960913	199730
WO 9710537	A3	19970424	WO 96US15664	A	19960913	199731
US 5659742	A	19970819	US 95528683	A	19950915	199739
US 5675788	A	19971007	US 95529233	A	19950915	199746
US 5742816	A	19980421	US 95529250	A	19950915	199823

Priority Applications (No Type Date): US 95529250 A 19950915; US 95528683 A 19950915; US 95529233 A 19950915

Cited Patents: No-SR.Pub; US 5241671; US 5404435; US 5404506; US 5524193

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 9710537	A2	E	90	G06F-000/00	
Designated States (National): AU CA CN JP MX NZ					
Designated States (Regional): AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE					
AU 9672026	A			G06F-019/00	Based on patent WO 9710537

US 5659742 A 43 G06F-017/30
US 5675788 A 44 G06F-017/30
US 5742816 A 43 G06F-017/30
WO 9710537 A3 G06F-000/00

Abstract (Basic): WO 9710537 A

The method for identifying textual documents and multimedia files involves storing a number of document and multimedia records each of which represent a document or multimedia file. The document records have associated text information fields, each of which represents text from one of the textual documents, and the multimedia records have multimedia information fields representing only digital video or audio information and associated text fields, each representing text associated with one of the multimedia information fields.

A single search query corresponding to the search topic is received pref in a natural language format, and an index database is searched in accordance with the single search query to simultaneously identify document records and multimedia records related to the single search query. A search result list having entries representing both textual documents and multimedia files related to the single search query is generated in accordance with the document records and the multimedia records identified by the index database search. Text or digital video or audio information corresponding to the search topic is retrieved by selecting entries from the search result list.

USE - Automated multi-user system for identifying and retrieving text and multi-media files from various publisher sources.

ADVANTAGE - Enables searching and retrieval of library or database to identify text documents and multimedia files relevant to query.

Dwg.4/12

Title Terms: IDENTIFY; TEXT; DOCUMENT; FILE; CORRESPOND; SEARCH; TOPIC;
ACCEPT; QUERY; RETURN; SINGLE; SEARCH; RESULT; LIST; TEXT; INFORMATION
Derwent Class: T01
International Patent Class (Main): G06F-000/00; **G06F-017/30** ; G06F-019/00
File Segment: EPI

14/5/39 (Item 31 from file: 350)

DIALOG(R) File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

009214394 **Image available**

WPI Acc No: 1992-341814/199242

XRPX Acc No: N92-260701

Bibliometric storage for full text document system - positioning by linear ordering to determine rank and storing rank as bibliometric property

Patent Assignee: KONINK PHILIPS ELECTRONICS NV (PHIG); PHILIPS

GLOEILAMPENFAB NV (PHIG); US PHILIPS CORP (PHIG)

Inventor: AALBERSBERG I J

Number of Countries: 004 Number of Patents: 005

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
EP 508519	A2	19921014	EP 92200891	A	19920330	199242 B
US 5293552	A	19940308	US 92860615	A	19920330	199410
EP 508519	A3	19930929	EP 92200891	A	19920330	199509
EP 508519	B1	20000531	EP 92200891	A	19920330	200031
DE 69231113	E	20000706	DE 631113	A	19920330	200039
			EP 92200891	A	19920330	

Priority Applications (No Type Date): EP 91200810 A 19910408

Cited Patents: No-SR.Pub; 3.Jnl.Ref; EP 304191

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
-----------	------	-----	----	----------	--------------

EP 508519	A2	E	11	G06F-015/401	
-----------	----	---	----	--------------	--

Designated States (Regional): DE FR GB

US 5293552	A	9	G06F-015/40	
------------	---	---	-------------	--

EP 508519	A3		G06F-015/401	
-----------	----	--	--------------	--

EP 508519 B1 E G-017/30
Designated States (Regional): DE FR GB
DE 69231113 E G06F-017/30 Based on patent EP 508519

Abstract (Basic): EP 508519 A

The bibliometric storage system accesses documents (20, 22, 24) containing a number of postings or elements such as noun, adverbs and the like. Each posting has a document identifier and an offset. The postings are in a linear order such as decreasing frequency of occurrence.

The information is then further compressed using pointers and tables to **remove redundant information**. This **compression** makes use of direct random addressability to reduce storage and stores all the postings for a particular document together.

ADVANTAGE - Allows easy expandability and updating while requiring roughly halve storage used by conventional systems.

Dwg.2/2

Title Terms: STORAGE; FULL; TEXT; DOCUMENT; SYSTEM; POSITION; LINEAR; ORDER
; DETERMINE; RANK; STORAGE; RANK; PROPERTIES

Derwent Class: T01

International Patent Class (Main): G06F-015/40; G06F-015/401; **G06F-017/30**

File Segment: EPI

?

?ds

Set	Items	Description
S1	29883	DUPLICAT? OR DUPE? ? OR COMMON? OR SAME? OR REDUNDAN? OR R- EPLICA? OR REPEAT? OR IDENTICAL? OR DOUBLE? OR ANALOGOUS? OR - SECOND OR PARALLEL? OR MULTIPLE?
S2	62627	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?.
S3	62648	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?
S4	3934	S1 (3N) S3
S5	26225	REMOV? OR DELETE? OR DELETING? OR ERASE? ? OR ERASING? OR - DEDUPE? ? OR CHECK? OR FIND???? OR IDENTIF? OR DETECT? OR C- OMPAR???? OR MATCH? OR PREVENT?
S6	188	S5(5N)S4
S7	28031	MERGE? OR MERGING? OR CONSOLIDAT? OR COMBIN? OR JOIN??? OR GROUP? OR COMPRESS? OR COMPOSITE?
S8	238	S7(5N)S4
S9	17	S6 AND S8
S10	16	S9 NOT PY>2001
S11	5456	MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL? OR CLINICAL?
S12	10	S6 AND S11
S13	7	S12 NOT S10

?show files

File 256:SoftBase:Reviews,Companies&Prods. 82-2004/Feb

(c)2004 Info.Sources Inc

?

?t. s10/5/4,6,8,9

10/5/4

DIALOG(R)File 256:SoftBase:Reviews,Companies&Prods.
(c)2004 Info.Sources Inc. All rts. reserv.

01165018 DOCUMENT TYPE: Product

PRODUCT NAME: iBaan for CRM (165018)

SSA Global Technologies Inc (241032)
500 W Madison #1600
Chicago, IL 60661 United States
TELEPHONE: (312) 258-6000

RECORD TYPE: Directory

CONTACT: Sales Department

Baan's iBaan for CRM is a customer relationship management (CRM) system that allows businesses to identify client requirements, improve order accuracy, and reduce service costs. The system **consolidates information** collected from **multiple** customer touchpoints, eliminating **redundant data** entry. iBaan for CRM also **identifies** profitable customers, allowing companies to prioritize and target marketing and sales activities. The system collects sales result, customer profile, qualified lead to completed sale ratio, and other information. iBaan for CRM provides companies with opportunity management, sales forecasting, automated quoting, automated proposal generation, and remote data synchronization features. The system supports online marketing and offline fulfillment operations.

DESCRIPTORS: CRM; Customer Service; Electronic Customer Service; Marketing Information

HARDWARE: Hardware Independent
OPERATING SYSTEM: Open Systems
PROGRAM LANGUAGES: Not Available
TYPE OF PRODUCT: Micro
POTENTIAL USERS: Cross Industry
PRICE: Available upon request

REVISION DATE: 20030915

10/5/6

DIALOG(R)File 256:SoftBase:Reviews,Companies&Prods.
(c)2004 Info.Sources Inc. All rts. reserv.

01058521 DOCUMENT TYPE: Product

PRODUCT NAME: Merge/Purge Plus (058521)

Group 1 Software Inc (400319)
4200 Parliament Pl
Lanham, MD 20706-1844 United States
TELEPHONE: (301) 731-2300

RECORD TYPE: Directory

CONTACT: Sales Department

Merge/Purge Plus, a product from Group 1 Software, detects duplicates in mailing lists. There are significant production costs involved in direct mail, and duplication can be costly, both in terms of wasted postage and wasted production units. **Merge /Purge Plus identifies** and eliminates **duplicate records** in a single file or across **multiple files** using a sophisticated **matching** algorithm. It can also recognize duplicate but

misspelled names. M/P Plus can also eliminate unwanted records. The user gains complete control over how the matching algorithm works, and can specify the algorithm as either loose, medium, tight, or equal for individual name matching. Users can also either include or exclude the first name as a requirement for duplicate detection, and can include apartment numbers as an extension of the street address, so that records are identified as duplicates only if the apartment numbers match. Merge/Purge Plus generates a full set of output files, including a mailing file fully cleaned of duplicates and purges, a duplicate names file, and an additional copy of the mailing file.

DESCRIPTORS: Data Quality; Direct Marketing; File Conversion; List Processing; Mailing Lists; Marketing Information

HARDWARE: IBM AS/400; IBM iSeries; IBM Mainframe; IBM PC & Compatibles; UNIX

OPERATING SYSTEM: Linux; MVS; OS/400; UNIX; VSE; Windows NT/2000; Windows XP

PROGRAM LANGUAGES: Not Available

TYPE OF PRODUCT: Mainframe; Mini; Micro; Workstation

POTENTIAL USERS: Marketing, Cross Industry, Direct Mail in U.S. and Canada

PRICE: Available upon request

REVISION DATE: 20031218

10/5/8

DIALOG(R) File 256:SoftBase:Reviews,Companies&Prods.

(c)2004 Info.Sources Inc. All rts. reserv.

00118040 DOCUMENT TYPE: Review

PRODUCT NAMES: Double Take Merge/Purge 4.1 Windows 9x & NT (663832)

TITLE: Low-cost Double Take keeps data honest

AUTHOR: Hollander, Geoffrey

SOURCE: InfoWorld, v21 n28 p56(1) Jul 12, 1999

ISSN: 0199-6649

HOME PAGE: <http://www.infoworld.com>

RECORD TYPE: Review

REVIEW TYPE: Review

GRADE: A

Peoplesmith Software's Double Take 4.1, a database cleaning and conditioning utility, gets very good marks overall, especially for the ability to process many files, file types, and sizes; several available problem-solving methods; versatile output choices; excellent processing performance rates; and Open Database Connectivity (ODBC) compliance. However, learning Double Take 4.1 takes some time, and the software does not yet support SQL Server 7.1. With Double Take 4.1, database administrators can streamline four pervasive data problems in companies that need squeaky clean database stability without getting into data warehousing. Testers could eliminate duplicate records from single and multiple file sets; merge multiple files into one output file; or combine the two tasks into one operation called merge /purge, which cleaned 2 million records per hour. Double Take 4.1 uses a matchcode, or a collection of name and address components, that works effectively to represent the one-of-a-kind identify of a complete record. Keys are created and then compared during the process that eliminates duplicate records. Double Take's best advantage is its ability to find near-duplicates with misspellings, typos, nicknames, and different or multiple addresses. The Matchcode Editor permits refining of keys that pinpoint these records.

PRICE: \$3000

COMPANY NAME: Peoplesmith Software (455458)

SPECIAL FEATURE: Charts

DESCRIPTORS: Database Utilities; File Conversion; IBM PC & Compatibles;
List Processing; Mailing Lists; Quality Assurance; Windows; Windows
NT/2000

REVISION DATE: 20030825

10/5/9

DIALOG(R) File 256:SoftBase:Reviews,Companies&Prods.

(c)2004 Info.Sources Inc. All rts. reserv.

00114542

DOCUMENT TYPE: Review

PRODUCT NAMES: MetaConnect for Database Applications (695556);
Meta-Directory 1.0 (741728)

TITLE: Netscape Teams with Isocor to Deliver Meta-Directory

AUTHOR: Sullivan, Thomas

SOURCE: ent, v4 n4 p18(2) Feb 17, 1999

ISSN: 1085-2395

HOME PAGE: <http://www.entmag.com>

RECORD TYPE: Review

REVIEW TYPE: Product Analysis

GRADE: Product Analysis, No Rating

Isocor's MetaConnect for Database Applications and Netscape Communications' Meta-Directory 1.0 will both be part of a new metadirectory technology to be jointly developed by the two vendors. Isocor is an Internet messaging and directory software developer. MetaConnect provides a **join** engine that **identifies duplicate** directory **records** in various directories, and automatically **merges** the information in one directory entry. The metadirectory platform will also use Isocor's connector technology to provide integration with information sources, which can include operating systems (OSs), telephony products, and database applications. Netscape is a licensee of Isocor's join technology, which is the basis of MetaConnect and Netscape's Meta-Directory 1.0 suite, which includes Netscape Directory Server, the metadirectory server, and some connectors. A central repository is required to build new Internet applications that allow transparent access to all of a company's data, says a senior product manager for Meta-Directory at Netscape. The suite will provide e-commerce applications with different directory information from human resource systems, databases, messaging systems, and network operating systems (NOSs). Important functions of the suite include the ability to move data from multiple directories into the Meta-Directory, to keep data current, and to join data using namespaces.

COMPANY NAME: Critical Path Inc (670413); Netscape Communications Corp
(592625)

SPECIAL FEATURE: Screen Layouts

DESCRIPTORS: Computer Resource Management; Network Administration; Network
Software

REVISION DATE: 20020630

?

?t s13/5/4

13/5/4

DIALOG(R) File 256:SoftBase:Reviews,Companies&Prods.
(c)2004 Info.Sources Inc. All rts. reserv.

01149195 DOCUMENT TYPE: Product

PRODUCT NAME: PERSIVAL (149195)

Columbia University (734403)
722 W 168th St #1030
New York, NY 10032 United States
TELEPHONE: (212) 305-3927

RECORD TYPE: Directory

CONTACT: Sales Department

The Personalized Retrieval and Summarization of Image, Video And Language (PERSIVAL) system, offered by Columbia University's NLP Group, provides clinicians and **patients** with access to a wide range of online **health** care information. PERSIVAL personalizes searches, summaries, and presentation elements for end users. It references **patient** records that are available at Columbia Presbyterian **Medical** Center (CPMC). It also allows users who are not CPMC **patients** to create new **patient** records. Referencing records, PERSIVAL automatically targets search results. The system identifies and summarizes multimedia content distributed across multiple local and online resources. It flags conflicting **information** and eliminates **duplicate** results. The system also **identifies** source types. Multimedia sources range from diagnostic test results to educational videos. PERSIVAL automatically summarizes video content.

DESCRIPTORS: Consumer **Health** Care; **Health** Care; Indexing; Information Retrieval; **Medical** Diagnosis; **Patient** Care; Personalization; Search Engines

HARDWARE: Hardware Independent
OPERATING SYSTEM: Open Systems
PROGRAM LANGUAGES: Not Available
TYPE OF PRODUCT: Micro
POTENTIAL USERS: Personalized Medical Information Retrieval
PRICE: Available upon request

REVISION DATE: 20030507

?

ds

Set	Items	Description
S1	19595	DUPLICAT? OR DUPE? ? OR COMMON? OR SAME? OR REDUNDAN? OR R- EPLICA? OR REPEAT? OR IDENTICAL? OR DOUBLE? OR ANALAGOUS? OR - SECOND OR PARALLEL?
S2	62627	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION - OR CITATION?.
S3	2607	REMOV? OR DELETE? OR DELETING? OR ERASE? ? OR ERASING? OR - DEDUPE? ?
S4	2085	S1(3N)S2
S5	25	S4 (3N)S3
S6	25980	MERGE? OR MERGING? OR CONSOLIDAT? OR COMBIN? OR JOIN??? OR GROUP?
S7	62	S6 (5N)S4
S8	2	S7 AND S5
S9	5396	MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL?
S10	0	S9 AND S5
S11	22	S5 NOT PY>2001
S12	0	(DUPE? OR DUPLICATE?) (3N) (CHECKER?)

?show files

File 256:SoftBase:Reviews,Companies&Prods. 82-2004/Feb
(c)2004 Info.Sources Inc

?t s11/5/3,4,6,9

11/5/3

DIALOG(R) File 256:SoftBase:Reviews,Companies&Prods.
(c)2004 Info.Sources Inc. All rts. reserv.

01169978 DOCUMENT TYPE: Product

PRODUCT NAME: My Labels 6.1 Standard & Premium (169978)

BVRP Software USA (643963)
1333 W 120th Ave #314
Westminster, CO 80234 United States
TELEPHONE: (720) 330-1400

RECORD TYPE: Directory

CONTACT: Sales Department

BVRP Software USA's My Labels, offered in Standard and Premium 6.1 editions, streamlines the creation and printing of labels. The system includes 10,000 clip art images. It also can import logos or other images. My Labels works with Avery (R), NEBS (R), PaperDirect (R), and other labels. Users can select from over 400 standard label sizes. Users also can define customized labels. The product imports information from a wide range of databases. It also can add POSTNET barcodes to labels. My Labels includes shrink-to-fit formatting and color adjustment features. It can print UPC-A, UPC-E, Code 39, Code 128, and other product barcodes. My Labels Premium 6.1 includes text alignment and rotation features. It includes 2.5 million clip art and photographic images. The program also allows users to **remove duplicate records** from imported **data**. Labels can be previewed before they are printed.

DESCRIPTORS: Barcoding; Database Utilities; Labels; Print Utilities

HARDWARE: IBM PC & Compatibles; Pentium
OPERATING SYSTEM: Windows; Windows NT/2000; Windows XP
PROGRAM LANGUAGES: Not Available
TYPE OF PRODUCT: Micro
POTENTIAL USERS: Cross Industry, Offices, Small Businesses
PRICE: \$19.95

TRAINING AVAILABLE: Technical support
OTHER REQUIREMENTS: 32MB RAM; Win 9x+ required
REVISION DATE: 20030921

11/5/4

DIALOG(R) File 256:SoftBase:Reviews,Companies&Prods.
(c)2004 Info.Sources Inc. All rts. reserv.

01134546 DOCUMENT TYPE: Product

PRODUCT NAME: Copernic Agent 6 Personal, Professional (134546)

Copernic Technologies Inc (640913)
360 Franquet #60
Ste-Foy, PQ G1N 4N3 Canada
TELEPHONE: (418) 527-0528

RECORD TYPE: Directory

CONTACT: Sales Department

Copernic Agent 6 Personal, Professional, and Basic from Copernic Technologies is a Web search engine that allows users to target Internet searches, organize results, access hidden data, create customized search

categories, and automatically remove results with broken links. Copernic Agent's parallel search features allow users to employ over 1,000 specialized search engines. The engines are organized into 120 categories. The program **removes duplicate records** from search results. Copernic Agent's management tools let users filter, group, and sort results; save pages; search within results; detect duplicate pages; view detailed search histories; refine search parameters; create custom categories; and export results in HTML, CSV, and other formats. The system integrates with Microsoft (TM) Internet Explorer and Microsoft Office. Users employ a toolbar to access search engines from the Microsoft programs. Copernic Agent includes context-sensitive help features. The Personal Edition adds features such as region filters, result-grouping tools, and Microsoft Office integration. The Professional Edition adds search updating, Web page analysis, page-language filters, date-modified filters, and much more.

DESCRIPTORS: Alerts; Front Ends; Internet Content Filters; Search Engines; Webmasters

HARDWARE: IBM PC & Compatibles

OPERATING SYSTEM: Internet Explorer; Open Systems; Windows NT/2000

PROGRAM LANGUAGES: Not Available

TYPE OF PRODUCT: Micro

POTENTIAL USERS: Cross Industry, Site Owners

PRICE: Available upon request

DOCUMENTATION AVAILABLE: Online documentation

REVISION DATE: 20030717

11/5/6

DIALOG(R) File 256:SoftBase:Reviews,Companies&Prods.

(c)2004 Info.Sources Inc. All rights reserved.

01121703 DOCUMENT TYPE: Product

PRODUCT NAME: Confort (121703)

Tripos Inc (481491)

1699 S Hanley Rd

St Louis, MO 63144-2913 United States

TELEPHONE: (314) 647-1099

RECORD TYPE: Directory

CONTACT: Sales Department

Tripos's Confort (TM) is a conformational analysis tool that examines drug-sized molecules. The product generates diverse conformers for receptor-ligand docking. It can be applied to entire structures or to user-defined substructures. Confort can search rings and acyclic rotatable bonds. The system **removes duplicate information**. Referencing user-defined number, energy, and diversity parameters, it outputs conformers. Confort can perform 3D database searches starting from multiple conformations. It can identify global minimum and low energy drug candidate conformations. The product handles the SMILES, UNITY hitlist/SLN, mol2, and MDL SDfile formats. Confort works with Tripos's UNITY (R) 3D search, DISCotech (TM) pharmacophore hypothesis generation, QSAR with CoMFA (R) multi-conformer 3D QSAR analysis, and FlexS (R) shape-based screening programs.

DESCRIPTORS: Bioinformatics; Molecular Modeling; Pharmaceuticals; Research & Development

HARDWARE: HP; IBM PC & Compatibles; Silicon Graphics; Sun; UNIX

OPERATING SYSTEM: HP-UX; IRIX; Oracle; Solaris; Windows; Windows NT/2000

PROGRAM LANGUAGES: Not Available

TYPE OF PRODUCT: Mini; Micro; Workstation

POTENTIAL USERS: Pharmaceuticals, Conformational Analysis
PRICE: Available upon request

REVISION DATE: 030206

11/5/9

DIALOG(R) File 256:SoftBase:Reviews,Companies&Prods.
(c)2004 Info.Sources Inc. All rts. reserv.

00119668 DOCUMENT TYPE: Review

PRODUCT NAMES: DB Assistant 2.0 Beta (775576)

TITLE: DB Assistant provides an ambulatory path through your data jungle
AUTHOR: Hollander, Geoffrey
SOURCE: InfoWorld, v21 n42 p68(1) Oct 18, 1999
ISSN: 0199-6649
HOMEPAGE: <http://www.infoworld.com>

RECORD TYPE: Review
REVIEW TYPE: Review
GRADE: A

Bloodhound Software's DB Assistant 2.0 beta, a data transformation software package, adroitly cleans data for use in data warehouses, data marts, or other decision support applications. Because clean, standardized data is required for such resources, DB Assistant's speed and efficiency, which result in much lower cleaning time than other products, are a significant advantage. Users can combine or nest tasks and reuse procedures. Processing speed is good and duplicate records can be weeded out. However, the merge module lacks grouping and cross-record summarization features, and the purge module cannot shift child records. DB Assistant 2.0 is more flexible, robust, and economical than competing products, including Smart DB Workbench and is more full-functioned than Data Junction or DBMS. Although DB Assistant 2.0 would be of great use to users of operating systems other than Windows, only Windows is supported. This release supports Object Linking and Embedding (OLE) for Databases, which in turn supports links to structured and unstructured data sources. DB Assistant 2.0 also works with many data, including text files, databases, spreadsheets, and legacy data. Matchcodes Administrator is used to **remove duplicate records**, by creating components that best represent the unique identity of each record, then comparing the results.

PRICE: \$4700

COMPANY NAME: Bloodhound Inc (669741)
SPECIAL FEATURE: Screen Layouts Charts
DESCRIPTORS: Data Marts; Data Warehouses; Decision Support Systems; File Conversion; Middleware
REVISION DATE: 20031125
?

ds

Set Items Description

S1 17263778 DUPLICAT? OR DUPE? ? OR COMMON? OR SAME? OR REDUNDAN? OR R-
EPLICA? OR REPEAT? OR IDENTICAL? OR DOUBLE? OR ANALOGOUS? OR -
SECOND OR PARALLEL? OR MULTIPLE?

S2 18205376 RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION
OR CITATION?.

S3 18227504 RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION
OR CITATION?

S4 14795806 REMOV? OR DELETE? OR DELETING? OR ERASE? ? OR ERASING? OR -
DEDUPE? ? OR CHECK? OR FIND???? OR IDENTIF? OR DETECT? OR C-
OMPAR???? OR MATCH? OR PREVENT?

S5 492535 S1(3N)S3

S6 16249 S5(5N)S4

S7 15616 S5(3N)(MERGE? OR MERGING? OR CONSOLIDAT? OR COMBIN? OR JOI-
N??? OR GROUP? OR COMPRESS? OR COMPOSITE?)

S8 538 S7 (S)S6

S9 43 S8 (S) (MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL? OR CL-
INICAL? OR EMR)

S10 36 RD (unique items)

S11 29 S10 NOT PY>2001

S12 29 S11 NOT PD>20010905

S13 94 S8 (S) (DATABASE? OR DATA()BASE? OR DATA(3N)(WAREHOUS? OR -
WARE()HOUS? OR MINE? ? OR MINING?) OR DATAMIN? OR DB OR DBS OR
RECORD?()MANAGE?)

S14 73 RD (unique items)

S15 67 S14 NOT PY>2001

S16 67 S15 NOT PY>20010905

S17 4 DUPE()CHECK? AND (ACT OR ARES)

?show files

File 813:PR Newswire 1987-1999/Apr 30
(c) 1999 PR Newswire Association Inc

File 16:Gale Group PROMT(R) 1990-2004/Mar 31
(c) 2004 The Gale Group

File 160:Gale Group PROMT(R) 1972-1989
(c) 1999 The Gale Group

File 634:San Jose Mercury Jun 1985-2004/Mar 30
(c) 2004 San Jose Mercury News

File 148:Gale Group Trade & Industry DB 1976-2004/Mar 30
(c)2004 The Gale Group

File 20:Dialog Global Reporter 1997-2004/Mar 31
(c) 2004 The Dialog Corp.

File 647:CMP Computer Fulltext 1988-2004/Mar W3
(c) 2004 CMP Media, LLC

File 674:Computer News Fulltext 1989-2004/Mar W3
(c) 2004 IDG Communications

File 149:TGG Health&Wellness DB(SM) 1976-2004/Mar W3
(c) 2004 The Gale Group

File 444:New England Journal of Med. 1985-2004/Mar W4
(c) 2004 Mass. Med. Soc.

?

5.27.04

12/3,K/5 (Item 1 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

13400668 SUPPLIER NUMBER: 72962643 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Streamline the Registration Process with EMPI. (Technology Information)
Roget, Melisa
Health Management Technology, 20, 5, 34
June, 1999
ISSN: 1074-4770 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 1676 LINE COUNT: 00146

... possible duplicate records were reviewed between January and September. More than 35,000 of these **records** were **identified** as **duplicates** and were **merged** - reducing the EMerge database to more than 1.6 million records. As the project grew to include multiple implementations, UAB decided additional resources were needed. The **healthcare** consulting firm, Insource Management Group, was enlisted to provide project management.

Phase 2: As of...

12/3,K/7 (Item 3 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

11114482 SUPPLIER NUMBER: 54858696 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Streamline the Registration Process with EMPI. (University of Alabama at Birmingham Health System's use of HIE's enterprise-wide master person index system, EMerge) (Company Operations)
Roget, Melisa
Health Management Technology, 20, 5, 34
June, 1999
ISSN: 1074-4770 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 1676 LINE COUNT: 00146

... possible duplicate records were reviewed between January and September. More than 35,000 of these **records** were **identified** as **duplicates** and were **merged** - reducing the EMerge database to more than 1.6 million records. As the project grew to include multiple implementations, UAB decided additional resources were needed. The **healthcare** consulting firm, Insource Management Group, was enlisted to provide project management.

Phase 2: As of...

12/3,K/8 (Item 4 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

10288817 SUPPLIER NUMBER: 20578239 (USE FORMAT 7 OR 9 FOR FULL TEXT)
How to find who. (master person indexes in health care systems integration) (includes a related articles on Matrix Rehabilitation's use of Datamedic's PMstation practice management solution and on TMC HealthCare's use of 3M's Healthcare Enterprise Management System) (Technology Information)
Cupito, Mary Carmen
Health Management Technology, v19, n6, p32(5)
May, 1998
ISSN: 1074-4770 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 3717 LINE COUNT: 00306

... expected to do a lot: receive patient or member information from many systems, assign unique **identifiers** to each individual, **merge records** pertaining to the **same** person, update its database, **remove**

'duplicate records , correlate a patient 's records throughout the enterprise, keep the source system's identifier or pointers to it...

12/3,K/13 (Item 9 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

05835963 SUPPLIER NUMBER: 12112227 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Cleanup and deduplication of an international bibliographic database.
Toney, Stephen R.
Information Technology and Libraries, v11, n1, p19(10)
March, 1992
ISSN: 0730-9295 LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT; ABSTRACT
WORD COUNT: 5595 LINE COUNT: 00443

... 2]June E. Slach, "Detection and Elimination of Duplicates from Multidatabase Searches," Bulletin of the **Medical** Library Association 73, no.3:235-37 (July 1985). [3]C. A. Giles and others, A Computerized Scheme for **Duplicate Checking** of Bibliographic **Data** Bases (NTIS CRNL/CSD-5, Aug. 1976). [4]C. A. Giles and others, An Experiment...

...CONF-76-1011-2, n.d.). [5]Thomas B. Hickey and David J. Rypka, "Automatic **Detection** of **Duplicate** Monographic **Records** ," Journal of Library Automation 12, no.2:125-42 (June 1979). [6]James L. Dolby...

...Retrieval," Journal of Library Automation 1, no.4:250-60 (Dec. 1968). [8]Pankaj Goyal, " **Duplicate Record Identification** in Bibliographic Databases," Information Systems 12, no.3:239-42 (1987). [9]E. J. Yannakoudakis...

12/3,K/14 (Item 1 from file: 20)
DIALOG(R)File 20:Dialog Global Reporter
(c) 2004 The Dialog Corp. All rts. reserv.

12133057 (USE FORMAT 7 OR 9 FOR FULLTEXT)
The American Association for Artificial Intelligence (AAAI) Awards the Best In Practical Innovations of Intelligent Systems Applications
PR NEWSWIRE
July 27, 2000
JOURNAL CODE: WPRW LANGUAGE: English RECORD TYPE: FULLTEXT
WORD COUNT: 677

(USE FORMAT 7 OR 9 FOR FULLTEXT)

... integrated illustration of the patients' medical condition(s), course of treatment and disposition while intelligently **deleting redundant information** , **combining** complementary **information** to create the integrated picture, and assessing the **medical** risks and service. LifeCode also has a "self-awareness" feature that recognizes its own limits...
?

?t s16/3,k/1,9,10,15,16,14,23,24,30,31,38,39,47,53,63

16/3,K/1 (Item 1 from file: 813)
DIALOG(R)File 813:PR Newswire
(c) 1999 PR Newswire Association Inc. All rts. reserv.

1405296 HSM024
**Netscape and ISOCOR Team to Deliver New Meta-Directory Technology; Netscape
Announces New Netscape(R) Meta-Directory 1.0**

DATE: January 18, 1999 08:29 EST WORD COUNT: 758

... join" engine that will be a key part of the combined meta-directory platform. The join engine identifies duplicate directory records in multiple directories and automatically combines the information into a single, more complete directory entry. The meta-directory platform will also...

...technology to provide integration with a variety of information sources, including operating systems, telephony products, database applications, and others.

"With business to business electronic commerce expected to grow to more than...

16/3,K/9 (Item 8 from file: 16)
DIALOG(R)File 16:Gale Group PROMT(R)
(c) 2004 The Gale Group. All rts. reserv.

06483716 Supplier Number: 55143940 (USE FORMAT 7 FOR FULLTEXT)
**Low-cost Double Take keeps data honest. (PeopleSmith's Double Take 4.1
backup software) (Software Review) (Evaluation)**
Hollander, Geoffrey
InfoWorld, v21, n28, p56
July 12, 1999
Language: English Record Type: Fulltext
Article Type: Evaluation
Document Type: Magazine/Journal; Trade
Word Count: 613

PeopleSmith's Double Take 4.1 provides an effective solution for database administrators to simplify four persistent data problems in companies that require immaculate database reliability without taking the plunge into high-end warehousing. During my tests, I was able to purge duplicate records from single and multiple file sets, merge several files into a single consolidated output file, or combine the two into a single operation known as a "merge/purge" at an impressive rate of 2 million records per hour.

Double Take uses a matchcode, or collection of name and address components that best represent the unique identity of a...
...BOTTOM LINE: VERY GOOD

Double Take 4.1

Summary: For database administrators who need to find and remove troublesome duplicate records or merge unique records from disparate databases, Double Take provides the most comprehensive, efficient, and reasonably priced answer - especially if you're not ready for a full-blown database warehousing solution.

Business Case: Duplicate records are costly and wreak havoc on every level -- from...

16/3,K/10 (Item 9 from file: 16)
DIALOG(R)File 16:Gale Group PROMT(R)
(c) 2004 The Gale Group. All rts. reserv.

06090516 Supplier Number: 53625395 (USE FORMAT 7 FOR FULLTEXT)

Netscape, Isocor Team On Meta Directory Technology 01/21/99

Newsbytes PM, pNA

Jan 21, 1999

Language: English Record Type: Fulltext

Document Type: Newswire; Trade

Word Count: 212

(USE FORMAT 7 FOR FULLTEXT)

TEXT:

...the companies say will be a "key part" of the combined meta- directory platform. The **join** engine reportedly **identifies duplicate** directory **records** in **multiple** directories and "automatically **combines** the information into a single, more complete directory entry." ISOCOR's connector technology will also...

...directory platform to provide "integration" with such information sources as operating systems, telephony products and **database** applications. Reported by Newsbytes News Network, <http://www.newsbytes.com>
. (19990121/Press Contact: ISOCOR Ltd...

16/3,K/15 (Item 14 from file: 16)

DIALOG(R)File 16:Gale Group PROMT(R)

(c) 2004 The Gale Group. All rts. reserv.

04358954 Supplier Number: 46392902 (USE FORMAT 7 FOR FULLTEXT)

New Software Product: "Data Clustering Engine"

News Release, pN/A

May 17, 1996

Language: English Record Type: Fulltext

Document Type: Magazine/Journal; Trade

Word Count: 293

(USE FORMAT 7 FOR FULLTEXT)

TEXT:

...The product can be useful in many applications for large organizations that have multiple, varied **databases** . It can be used to help eliminate **duplicate records** or **identify** various relationships between records. It can also be used for "**data mining**" purposes for customer information, marketing or fraud investigation systems. The product can also be used as a tool to assist in **combining** multiple **files** into one "**data warehouse**", both to set up the warehouse for the first time and also as an on...

...any other information that the user chooses. The product uses it's own high performance **database** which can access data from any of the popular **databases** . The product has its own Application Programming Interface (API) which application programs can use to...

16/3,K/16 (Item 15 from file: 16)

DIALOG(R)File 16:Gale Group PROMT(R)

(c) 2004 The Gale Group. All rts. reserv.

03340582 Supplier Number: 44622834 (USE FORMAT 7 FOR FULLTEXT)

Access Database Manager Comes Of Age

Open Systems Today, p85

April 25, 1994

Language: English Record Type: Fulltext

Document Type: Newsletter; Trade

Word Count: 2433

... query's data will come. These tables can be attached to either local or remote **databases** , and one query can **combine data** from **multiple databases** . You then define or **remove** any default relationships between these tables, as needed. Then select each field you want to...

16/3,K/18 (Item 17 from file: 16)
DIALOG(R)File 16:Gale Group PROMT(R)
(c) 2004 The Gale Group. All rts. reserv.

02637258 Supplier Number: 43510891 (USE FORMAT 7 FOR FULLTEXT)
Software Converts Legacy To Other Formats 12/07/92
Newsbytes, pN/A
Dec 7, 1992
Language: English Record Type: Fulltext
Document Type: Newswire; General Trade
Word Count: 659

... minutes, the executive VP said. Without performing any programming, the user can specify conditional queries, **merge data** from **multiple** sources, and **check data** integrity, operations previously possible only through handwritten programs. The Data Conversion Tool also produces detailed documentation of the conversion and the inter- **database** relationships.

The Extract Libraries, a product still in the pilot phase, consists of packaged definitions...

16/3,K/14 (Item 13 from file: 16)
DIALOG(R)File 16:Gale Group PROMT(R)
(c) 2004 The Gale Group. All rts. reserv.

04402086 Supplier Number: 46457689 (USE FORMAT 7 FOR FULLTEXT)
Apertus Announces New Relationship with Postalsoft, Extends Relationship with Object Design Flagship Product Enterprise/Integrator Drives Addition of "Best-of-Breed" Partners.
Business Wire, p06111304
June 11, 1996
Language: English Record Type: Fulltext
Document Type: Newswire; Trade
Word Count: 548

... oriented development tool for creating applications that transform, cleanse, merge, and synchronize data from heterogeneous **databases**. More than a simple data transformation tool, Enterprise/Integrator allows users to graphically develop solutions for complex data quality problems, such as **identifying** and **consolidating** logically **redundant data** and resolving value conflicts. The product is available on Sun SPARC, HP 9000 Series, and...

16/3,K/23 (Item 5 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

12343398 SUPPLIER NUMBER: 21280088 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Measuring the performance of a merging algorithm: mismatches, missed-matches, and overlap in Israel's Union List. (Union List of Monographs)
Meir, Daniel D.; Laszinger, Susan S.
Information Technology and Libraries, v17, n3, p116(8)
Sept, 1998
ISSN: 0730-9295 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 4207 LINE COUNT: 00395

... Merge and Not to Merge," 216.
(5.) Martha E. Williams and Keith D. MacLaury, "Automatic **Merging** of Monographic **Data Bases -- Identification of Duplicate Records in Multiple Files** : The IUCS Scheme," Journal of Library Automation 12 (June 1979): 166.
(6.) Stephen R. Toney...

16/3,K/24 (Item 6 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

12247377 SUPPLIER NUMBER: 62827319 (USE FORMAT 7 OR 9 FOR FULL TEXT)
New Ovid Release Enhances Searching.
EContent, 23, 3, 8
June, 2000
ISSN: 1525-2531 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 88 LINE COUNT: 00011

TEXT:

...and Deduping, which lets users combine multiple Ovid data bases,
search them simultaneously, and automatically **remove duplicate records**
from search results. (800/950-2035; 212/563-3006; <http://www.ovid.com>)

16/3,K/30 (Item 12 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

10782870 SUPPLIER NUMBER: 53682915 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Cumulative Contents of the Journal of Library Automation,
1968-1981. (Directory)
Information Technology and Libraries, 17, 4, 233(1)
Dec, 1998
DOCUMENT TYPE: Directory ISSN: 0730-9295 LANGUAGE: English
RECORD TYPE: Fulltext
WORD COUNT: 5731 LINE COUNT: 00499

... of Fixed-Length Keys Derived from Title Strings" (p.143-55), Keith
D. MacLaury
"Automatic **Merging** of Monographic Data Bases -- Identification
of **Duplicate Records** in Multiple Files : The IUCS Scheme"
(p.156-68), Martha E. Williams and Keith D. MacLaury
Communications: "Maximum..."

16/3,K/31 (Item 13 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

10434034 SUPPLIER NUMBER: 21082541 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Technology Closeup: Main Entry Point For Contacts -- Support for vCard is
key. (contact manager software that support the vCard standard)
(Technology Information)
Correia, Edward J.
Computer Reseller News, n805, p99(1)
August 31, 1998
ISSN: 0893-8377 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 322 LINE COUNT: 00027

... the others can duplicate.
The next best supporter of the standard was ACT, which can **merge**
vCards into **database records** with **dupe checking**. Neither the
Organizer 5.0 beta nor GoldMine 4.0 features vCard duplicate checking; both
...

16/3,K/38 (Item 20 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

08061585 SUPPLIER NUMBER: 17169933 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Apertus Announces Development Tool for Enterprise-wide Data Integration

Enterprise/Integrator s in Building the Data Warehouse

Business Wire, p8141136

August 14, 1995

LANGUAGE: English RECORD TYPE: Fulltext

WORD COUNT: 426 LINE COUNT: 00053

TEXT:

...transformation tool, it allows the development of solutions for complex data integration issues that include **identifying** and **consolidating** logically **redundant data** such as names and addresses. It also resolves value conflicts such as **identifying** a correct address from **multiple**, inconsistent **records**. (See August 8 Dow Jones article, "**Data Warehouses Seen As Big Opportunity for Tech Cos.**")...

16/3,K/39 (Item 21 from file: 148)

DIALOG(R)File 148:Gale Group Trade & Industry DB

(c)2004 The Gale Group. All rts. reserv.

07544745 SUPPLIER NUMBER: 15787181 (USE FORMAT 7 OR 9 FOR FULL TEXT)

To merge and not to merge: Israel's Union List of Monographs in the context of merging algorithms.

Lazinger, Susan S.

Information Technology and Libraries, v13, n3, p213(7)

Sept, 1994

ISSN: 0730-9295 LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT; ABSTRACT

WORD COUNT: 5819 LINE COUNT: 00472

... cleanup and deduplication" algorithm described by Toney, the first pass inspected the entire database and **grouped** the **records** into "pools." The **second** pass examined each pool with a more precise algorithm to find duplicates. The advantage to...together records derived from a standard LC cataloging in a minimum of processing time. A **second** phase processed **records** that didn't **match** on the minimum set of data elements through the remainder of the algorithm, comparing the...

...of Illinois algorithm consisted of a first step that generated a key to partition the **database** into many small sets of possible duplicates based on the last two digits of the...

...with its very tight matching criteria, state that "the ... program committed very few errors in **matching** different **records** but missed many **duplicates** ." (26) Thus, in one sentence they pinpoint the major problem which must be faced in...of Library Automation 12 (June 1979): 131; Martha E. Williams and Keith D. MacLaury, "Automatic **Merging** of Monographic **Data Bases -- Identification of Duplicate Records in Multiple Files** : The IUCS Scheme," Journal of Library Automation 12 (June 1979): 159-60.

(17.)Coyle, "Record...

16/3,K/47 (Item 29 from file: 148)

DIALOG(R)File 148:Gale Group Trade & Industry DB

(c)2004 The Gale Group. All rts. reserv.

05835963 SUPPLIER NUMBER: 12112227 (USE FORMAT 7 OR 9 FOR FULL TEXT)

Cleanup and deduplication of an international bibliographic database.

Toney, Stephen R.

Information Technology and Libraries, v11, n1, p19(10)

March, 1992

ISSN: 0730-9295 LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT; ABSTRACT

WORD COUNT: 5595 LINE COUNT: 00443

... s own. [Tabular Data Omitted]

REFERENCES

[1]Martha E. Williams and Keith D. MacLaury, "Automatic **Merging** of Monographic **Data Bases -- Identification of Duplicate Records in Multiple Files** : The IUCS Scheme," Journal of Library Automation 12,

no.2:156-68 (June 1979).

...3:235-37 (July 1985). [3]C. A. Giles and others, A Computerized Scheme for **Duplicate** Checking of Bibliographic **Data Bases** (NTIS CRNL/CSD-5, Aug. 1976). [4]C. A. Giles and others, An Experiment in...

...CONF-76-1011-2, n.d.). [5]Thomas B. Hickey and David J. Rypka, "Automatic **Detection** of **Duplicate** Monographic **Records**," Journal of Library Automation 12, no.2:125-42 (June 1979). [6]James L. Dolby...

...Retrieval," Journal of Library Automation 1, no.4:250-60 (Dec. 1968). [8]Pankaj Goyal, " **Duplicate Record Identification** in Bibliographic **Databases**," Information Systems 12, no.3:239-42 (1987). [9]E. J. Yannakoudakis and others, "Matching of Citations Between Non-Standardized **Databases**," Journal of the American Society for Information Science 41, no.8:599-610 (1990). [10...

16/3,K/53 (Item 35 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

04779474 SUPPLIER NUMBER: 09065495 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Detecting duplicates: a searcher's dream come true.

Miller, Carmen

Online, v14, n4, p27(8)

July, 1990

ISSN: 0146-5422 LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT

WORD COUNT: 4341 LINE COUNT: 00342

TEXT:

...begins to search multiple databases with overlapping sources, the need for a way to sort **merged** output, **identify** the **duplicates**, and retain the **records** one chooses becomes imperative. Duplicate **detection** is a longawaited, major system enhancement which allows searchers to do all of these things.

16/3,K/63 (Item 1 from file: 647)
DIALOG(R)File 647:CMP Computer Fulltext
(c) 2004 CMP Media, LLC. All rts. reserv.

01170679 CMP ACCESSION NUMBER: CRN19980831S0096

Technology Closeup: Main Entry Point For Contacts - Support for vCard is key (Contact Management)

Edward J. Correia

COMPUTER RESELLER NEWS, 1998, n 805, PG99

PUBLICATION DATE: 980831

JOURNAL CODE: CRN LANGUAGE: English

RECORD TYPE: Fulltext

SECTION HEADING: CRN Test Center

WORD COUNT: 302

... the others can duplicate.

The next best supporter of the standard was ACT, which can **merge** vCards into **database records** with **dupe checking**. Neither the Organizer 5.0 beta nor GoldMine 4.0 features vCard duplicate checking; both...

?

?ds

Set	Items	Description
S1	7040004	DUPLICAT? OR DUPE? ? OR COMMON? OR SAME? OR REDUNDAN? OR R- EPLICA? OR REPEAT? OR IDENTICAL? OR DOUBLE? OR ANALOGOUS? OR - SECOND OR PARALLEL? OR MULTIPLE?
S2	10979839	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?.
S3	10990032	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?
S4	6119049	REMOV? OR DELETE? OR DELETING? OR ERASE? ? OR ERASING? OR - DEDUPE? ? OR CHECK? OR FIND???? OR IDENTIF? OR DETECT? OR C- OMPAR???? OR MATCH? OR PREVENT?
S5	330820	S1(3N)S3
S6	11940	S5(5N)S4
S7	9102853	MERGE? OR MERGING? OR CONSOLIDAT? OR COMBIN? OR JOIN??? OR GROUP? OR COMPRESS? OR COMPOSITE?
S8	16089	S7(5N)S5
S9	882	S6 AND S8
S10	2755530	MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL? OR CLINICAL?
S11	177	S10 AND S9
S12	2755530	MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL? OR CLINICAL?
S13	177	S11 AND S12
S14	176	S11(S)S12
S15	159	RD (unique items)
S16	131	S15 NOT PY>2001
S17	130	S16 NOT PD>20010905
S18	44	S17(S) (DATABASE? OR DATA()BASE? OR DATA(3N) (WAREHOUS? OR - WARE()HOUS? OR MINE? ? OR MINING?) OR DATAMIN? OR DB OR DBS OR RECORD? ()MANAGE?)
S19	2	S17 (S) (EMR OR ELECTRONIC()MEDICAL()RECORD?)
S20	3	S17 AND (EMR OR ELECTRONIC()MEDICAL()RECORD?)
S21	3	S20 AND S6
S22	43	S18 NOT S21
S23	108	S9 (10N) (DATE? OR TIME? OR CONTENT?)
S24	102	RD (unique items)
S25	85	S24 NOT PY>2001
S26	1	DUPE()CHECK? AND (ACT OR ARES)

?show files

File 15:ABI/Inform(R) 1971-2004/Mar 30
(c) 2004 ProQuest Info&Learning

File 9:Business & Industry(R) Jul/1994-2004/Mar 30
(c) 2004 The Gale Group

File 610:Business Wire 1999-2004/Mar 31
(c) 2004 Business Wire.

File 810:Business Wire 1986-1999/Feb 28
(c) 1999 Business Wire

File 275:Gale Group Computer DB(TM) 1983-2004/Mar 31
(c) 2004 The Gale Group

File 476:Financial Times Fulltext 1982-2004/Mar 31
(c) 2004 Financial Times Ltd

File 624:McGraw-Hill Publications 1985-2004/Mar 30
(c) 2004 McGraw-Hill Co. Inc

File 621:Gale Group New Prod.Annou.(R) 1985-2004/Mar 31
(c) 2004 The Gale Group

File 636:Gale Group Newsletter DB(TM) 1987-2004/Mar 31
(c) 2004 The Gale Group

File 613:PR Newswire 1999-2004/Mar 31
(c) 2004 PR Newswire Association Inc

File 618:Xinhua News 1999-2004/Mar 30
(c) 2004 Xinhua News via Comtex

?

21/3,K/1 (Item 1 from file: 15)
DIALOG(R) File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

02521370 116356554

Issues in the multi-disciplinary assessment of healthcare information systems

Heathfield, Heather; Hudson, Peter; Kay, Stephen; Mackay, Lesley; Marley, Tom; Nicholson, Lorraine; Peel, Victor; Roberts, Ruth; Williams, John
Information Technology & People v12n3 PP: 253-275 1999
ISSN: 0959-3845 JRNL CODE: OTP
WORD COUNT: 9650

Issues in the multi-disciplinary assessment of healthcare information systems

...ABSTRACT: considers the problems of a multi-disciplinary team working together to understand and evaluate a **healthcare** information system, which itself is situated in a complex organisational and political environment. A general...

... faced by evaluators of such systems is provided. A specific evaluation project is described (Electronic **Patient** Records in the UK National **Health** Service), and an account of the evaluation process as it occurred is given. Some of...

TEXT: Heather Heathfield : Computing Department, Manchester Metropolitan University, Manchester, UK

Peter Hudson: **Health** Services Management Unit, University of Manchester, Manchester, UK

Stephen Kay: Department of Computing, University of...

... Tom Marley: Department of Computing, University of Manchester, Manchester, UK

Lorraine Nicholson and Victor Peel: **Health** Services Management Unit, University of Manchester, Manchester, UK

Ruth Roberts and John Williams: School of Postgraduate Studies in **Medical** and **Health** Care, Swansea, UK

Introduction

This paper examines some of the important, and often difficult issues...

...source of data it provides.

The subject of this evaluation project is that of Electronic **Patient** Records (EPRs) which are a relatively new and unexplored application domain. The first section of...

...by the team, about the process of working as a multi-disciplinary team.

Background

Electronic **patient** record systems

The use of information technology in **healthcare** is not new. The first digital computers were developed in the 1940s, and by the 1960s **Hospital** Information Systems (HIS) began to emerge. HIS are intended primarily as a means of communication, providing order-entry and results-reporting functionality. Whilst they may store some basic **patient** details, they do not contain detailed **clinical** information and are not intended to directly support **clinical** care. HIS are routinely used in many **hospitals** in both the UK and USA, as well as developing countries.

Given the apparent success of HIS, attention turned towards the computerisation of **patient** records. Paper-based records have many problems associated with their use, for example availability, illegibility ...

... EPRs as a solution to these problems, and likely to lead to significant improvements in **healthcare**. Research into the design and development of EPRs began in the 1970s. Three important early systems were developed in the USA, namely COSTAR (Barnett et al., 1979), the Regenstrief **Medical** Record System (Pryor et al., 1983) and The Problem Orientated **Medical** Record System (Weed, 1969).

Today, EPRs are an important research area in **health** informatics, as can be seen by the numerous EPR projects, conferences devoted to the topic...

... In 1991 the USA Institute of Medicine published a committee report titled The Computer-based **Patient** Record: An Essential Technology for **Patient** Care which stated that computer-based **patient** record systems have a unique potential to improve **patient** care (Dick and Steen, 1991). This report has been extremely influential in setting out the requirements of computer-based **patient** record systems. Work on EPRs is progressing in the USA (e.g. Carpenter, 1994; Churgin...

...and Steen, 1995) and Europe (e.g. van der Lei et al., 1993).
Evaluating electronic **patient** records

Evaluations of **healthcare** applications first received serious attention in the 1970s (e.g. de Dombal et al., 1974...

...and "how" of evaluation (e.g. Wyatt and Spiegelhalter, 1990).

As with any new intervention, **healthcare** applications do not exist independently of the economic realities that shape their design and influence the potential for their successful use in routine settings. The majority of **healthcare** systems were conceived as research projects, funded by government or private grants. The ultimate transmission...

... had expired. Thus, the ability to demonstrate economic benefits is seen as essential by many **healthcare** professionals. However, there are other recognised potential benefits of IT, such as improved quality of **patient** care, brought about, by example, through better access to more accurate information (Detmer and Steen, 1995).

Economic analyses can be difficult for large **healthcare** applications of IT, and there are severe problems in applying the results of such analyses ... Van der Loo (1995) examined the type of study design employed in the evaluation of **healthcare** information systems (1974-1995). This study set out a "gold standard" of evaluation which comprises an economic analysis coupled with a randomised controlled **clinical** trial (RCT), and scored the evaluation studies examined against this standard. Of the 108 studies...

... studies and there is a strong body of opinion which advocates the need for all **clinical** systems to undergo RCTs. This debate is still highly topical as illustrated by two recent papers in the British **Medical** Journal (Donaldson, 1996; Locke, 1996).

Donabedian's classic work of the 1960s, recognised that the quality of **health** services can be measured along three dimensions: structure (or resources), process and outcomes (Donabedian, 1966...

... systems focused on process and evaluations of treatment systems focused on outcomes, systems to support **clinical** care were generally examined in terms of structure.

The application of RCTs to EPRs is...

... recipients from which to randomise, introducing randomness into the implementation of whole or part of **hospital** systems is impossible. Each institution will have many different characteristics which it would be impossible...

... into, virtually, a "national programme". Even randomisation within a particular institution (for example, randomly allocating **patients** to a clinician who uses a particular EPR system and one who does not) will...

... an objective, when in reality this is an intermediate benefit that may produce benefits in **patient** care or the organisational setting. The second point highlighted by this review was that most...

... recently, there is a growing awareness amongst those involved in the development and implementation of **clinical** systems, that social and organisational issues are at least of equal importance as technical issues
...

... promote understanding of why such problems occurred (Rotman et al., 1996).

Recently the Computer-based **Patient** Record Institute of the USA published a discussion document on system evaluation. It gave support...

... comparisons difficult. To complicate matters the term EPR is often used interchangeably with Computer-based **Patient** Record and also **Clinical** Workstation.

- The determination of the evaluation questions. Given the plethora of possible questions how does... the early 1970s with a few isolated but important local initiatives which sought to establish **hospital** wide **clinical** information systems, e.g. at the London **Hospital** (Barber et al., 1994). In the 1970s there was an attempt to build a national standard **hospital** information system. However, following an evaluation and review, this project was terminated in 1980. In 1989 the NHS launched the **Hospital** Information Support Systems (HISS) Project. This funded several pilot sites in the development of electronic communication systems to provide ward order-entry and results-reporting for individual **patients**. In addition to results, these systems store some basic **patient** identification details and an administrative, computerised index to those **patients**. They do not generally contain detailed **clinical** information and are not intended to directly support **clinical** care. Their role is to assist in the administration of care, the reimbursement of costs...

... the UK is currently estimated to be in the region of approximately 20 district general **hospitals**, from a total of 350 which each have more than 300 beds.

In 1994 the...

...through its Information Management Group (IMG), initiated two programmes of work which sought to improve **patient** records using information technology. The Electronic **Patient** Record (EPR) project is a three year research and development programme to assist clinicians in acute **hospitals** to provide better care to **patients** through the use of electronic **patient** record systems (NHS Executive, 1994a). The **Clinical** Workstation (ICWS) project is a two year project concerned with providing interface facilities to an...

... The PIDs were structured along the lines of sub-projects, for example document archiving or **clinical** terms.

Burton **Hospitals** NHS Trust and Wirral **Hospital** NHS Trust are the demonstrator sites for the EPR project. Both these sites had established...

...additional funds to extend the functionality of these systems to provide more comprehensive and richer **clinical** data for routine **clinical** care. The demonstrator sites for the ICWS project are Winchester and Eastleigh

HealthCare NHS Trusts **Bethlem** and **Maudsley NHS Trust** and **Horizon NHS Trust**.

Despite the fact that there has been almost 30 years of experience with **hospital** -based **clinical** information systems in the NHS, published studies of system evaluation are scarce. Those that do...

... on both system costs and benefits. Given the high estimated level of investment in UK **hospital** systems (Pounds 220million per annum) the Audit Commission (1995a, 1995b) has stated that there have...

... sites themselves, other sites involved or about to embark on EPR/ICWS development, manufacturers of **clinical** systems, the EPR and ICWS Programme Boards and the Research and Development Directorate (RDD) of...

... setup after previous HIS evaluation work (Welsh Office, 1993, 1996) identified key evaluation questions concerning **clinical**, organisation, technical, financial and managerial issues. PROBE provides guidance for the NHS on evaluating large...

... both technical and social factors, using quantitative and qualitative methods, is currently considered important in **healthcare**. The Centre for **Health Informatics** at the **Health Services Management Group** of the University of Manchester instigated the formation of a multi-disciplinary team which included clinicians, social scientists, computer scientists, **health** service management experts and economists. They provide a combination of experience of the **clinical** environment, **healthcare** organisation, **medical** informatics technologies, **patient** records and the NHS organisation. They also have experience of ...projects. The projects had already started preventing any straightforward before and after comparisons. From a **clinical** perspective the RDD were keen to demonstrate improvements in **patient** care, as documented by **patient** outcomes. The problem lies in the fact that such documentation is not currently available and...

...underlying team principles:

- In this particular project we cannot distinguish (for evaluation purposes) between a **Hospital** Information System (HIS), EPR and ICWS. The individual EPR/ICWS sites draw on the data...

... combinations of paper-based and electronic data which together form the concept of a "total **patient** record". Our evaluation must explore the use and consequences of these combinations.

- A successful evaluation... defined by the team, are listed below. They include a wide range of perspectives, including **clinical**, technical, social, cultural, organisational, and managerial, reflecting the different backgrounds and perspectives of the team.

- What is the impact of the technology on **clinical** management? This will be considered at three levels: individual **patient** care, management of services and resource management.

- What is the impact on the roles, the...

...at the implementation sites?

- Can the costs and benefits of such developments/technologies be valued?

- **Patient** record systems and technologies: how useful and usable are they?

- What is the relationship between...

... clinicians and the system) rather than as a technique to gain wider understanding of the **hospital** context. However, this overlap in methods between different groups initially caused some problems, as different...

... study. It would be very difficult to separate the effect of these individual technologies on **clinical** practice. Furthermore, the recipients of the evaluation results may not perceive such an arbitrary distinction but instead makes the concept of a "total **patient** record" central to the framework. This allows us to examine the relationship between paper and...

... employed and locations in which the record is stored. Using the concept of a "total **patient** record" also enables us to highlight specific relationships which are more pertinent to an EPR...

...listed above. A simplified example is given below.

- Take research question one of the list. **Patient** record systems and technologies: how useful and usable are they? One of the specific criteria ...

... selected sites. Methods used include observations, structured and semi-structured interviews, questionnaires and examination of **patient** records. Many different personnel at the sites have been included in the study: nursing staff, **medical** staff, managers, technical staff, and sub-project leaders. A great deal of effort was needed...

... setup the site visits at times that were convenient for both the team members and **hospital** staff, and in a manner that caused as little disruption as possible to the sites...because they made our work easier.

- The RDD pressed for documentation which demonstrated improvements in **patient** care. This was inevitably difficult, if not impossible, because of the evaluation commencing well after...

...date, the interface between HIS and EPR being extremely blurred, and the paucity of analysable **clinical** data which failed to provide a baseline for comparison.

As in the scoping study, the...as a descriptive story to the sites, whereas the computer scientists attempted to abstract their **findings** into areas such as **duplication** of **information** or interface issues. The aim of these sessions was to check that the site agreed...of the project, a subset of team members produced a progress report for a major **healthcare** and IT conference. This took the form of a short paper, which, the authors thought ...

...References

1. Anderson, J.G., Aydin, C.E. and Jay, S.J. (Eds) (1994), **Evaluating Health Care Information Systems, Methods and Applications**, (1st ed.), Sage Publications, London.
2. Audit Commission (1987), **Use of Operating Theatres in the National Health Service**, HMSO, London.
3. Audit Commission (1995a), **For Your Information - A Study of Information Management and Systems in the Acute Hospital**, HMSO, London.
4. Audit Commission (1995b), **Setting the Records Straight - A Study of Hospital Medical Records**, HMSO, London.
5. Barber, B., Abbot, W. and Peel, V. (1994), **Information Management in Health Care**, IHSM/Longmans, London.
6. Barnett, G.O., ... M.S., Van Deusen, F.R. and Greenlie, J.K. (1979), "COSTAR - a computer-based **medical** information system for ambulatory care", IEEE, Vol. 67, pp. 1222-37.
7. Boloix, G. and...

...framework", ACM, Vol. 28 No. 12, pp. 17-26.

8. Carpenter, P.C. (1994), "The **electronic medical record** : perspective from Mayo Clinic", Int J Biomed Comp, Vol. 34, pp. 159-71.

9. Churgin, P.G. (1994), 'Introduction of an automated **medical record** at an HMO clinic", MD Computing, Vol. 11, pp. 293-300.

10. Davies, D. and Nicholas, E. (1995), "CPR recognition program", in Steen, E.B. (Ed.), Proceedings, Computer-based **Patient Record Institute Inc.**, Bethesda, MD.

11. de Dombal, F.T., Leaper, D.J., Staniland, J...

... 12. Detmer, D.E. and Steen, E.B. (1995), "Countdown to 2001: the computer-based **patient record** after the institute of medicine report", IMIA Yearbook of **Medical Informatics** 1995; F.K. Schattauer Verlagsgesellschaft mbH, Stuttgart, pp. 55-60.

13. Dick, R.S. and Steen, B. (Eds) (1991), The Computer-based **Patient Record**, An Essential Technology for **Healthcare**, National Academy Press, Washington, DC.

14. Donabedian, A. (1966, "Evaluating the quality of **medical care**", Milbank Memorial Fund Quarterly, Vol. 44, p. 166.

15. Donaldson, L.J. (1996), "From...

... Heathfield, H.A. et al. (1996), "Evaluating EPR and CWS - integrating technical and organisational factors", **Health Care Computing** 96, BJHC Limited, Weybridge.

18. HISS Evaluation Programme (1996), First Report of PROBE...

...Leeds, March.

19. Kaplan, B. (1995), "A model comprehensive evaluation plan for complex information systems: **Clinical** imaging systems as an example", Conference Proceedings, Information Technology Conference 1995, Henley Management College, Henley on Thames.

20. Locke, C. (1996), "What value do computers provide to NHS **hospitals** ", BMJ, 312, pp. 1407-10.

21. Lorenzi, N.M. (1996), Integrated Academic Information Management System, To Manage the Knowledge of Medicine More Effectively, University of Cincinnati **Medical Center Report**, Cincinnati, OH.

22. NHS Executive (1994a), Electronic **Patient Record** Initiation Document, NHSE, Leeds.

23. NHS Executive (1994b), Integrated **Clinical Workstation** Initiation Document, NHSE, Leeds.

24. Nygren, E. and Henriksson, P. (1992), "Reading the **medical record**. I", Computer Methods and Programs in Biomedicine, Vol. 39, pp. 1-12.

25. Pope...

... N. (1995), "Reaching the parts other methods cannot reach: an introduction to qualitative methods in **health and health services research**", BMJ, Vol. 311, pp. 42-5.

26. Protti, D.J. and Haskell, A.R. (1992), "Managing information in **hospitals** : 60 per cent social, 40 per cent technical", in: Ehlers, C. Bakker, A. Bryant, J. Hammond, W. (Eds), Proceedings of IMIA Working Conference on Trends in **Hospital Information Systems**, North Holland Publishings, Amsterdam, April 1992.

27. Pryor, T.A. et al (1983), "The HELP system", *Journal of Medical Systems*, Vol. 7, p. 87.

28. Rotman, B.L., Sullivan, A.N., McDonald, T.W...

... Implementation barriers to overcome evaluation", *JAMIA*, Vol. 3, pp. 340-8.

29. Weed, D. (1969), *Medical Records, Medical Education and Patient Care*, The Press of Case Western Reserve University, Cleveland, OH.

30. Welsh Office (1993), *Evaluation...*

...Information Systems and Benefits, NHS Cymru Wales.

32. Wyatt, J. and Spiegelhalter, D. (1990), "Evaluating **medical** expert systems: what to test and how?", *Meth Inform*, Vol. 15, pp. 205-17.

33...

... M., Boon, W.M. and van Bommel, J.H. (1993), "The introduction of computer-based **patient** records in the Netherlands", *Ann Intern Med*, Vol. 19, pp. 1036-41.

34. Van der...

COMPANY NAMES:

National **Health** Service-UK...

...DESCRIPTORS: **Health** informatics

21/3,K/2 (Item 2 from file: 15)

DIALOG(R)File 15:ABI/Inform(R)

(c) 2004 ProQuest Info&Learning. All rts. reserv.

01477652 01-28640

An international collaboratory based on virtual patient records

Kilman, David G; Forslund, David W

Communications of the ACM v40n8 PP: 110-117 Aug 1997

ISSN: 0001-0782 JRNL CODE: ACM

WORD COUNT: 3979

An international collaboratory based on virtual patient records

ABSTRACT: Virtual **patient** records will impact **health** care internationally the way the Internet has transformed global communications. Virtual **patient** records are virtual in the sense they provide a view of data possibly configured differently...

... locations, but united together into a common format for viewing at the required time. Virtual **patient** records provide a user with virtual access to data possibly scattered around the world. The virtual **patient** record should form the basis for a collaboratory environment in which multiple physicians, and ultimately the **patient**, should be able to engage in an interactive electronic discussion. International implications and standardization issues are discussed in detail. TeleMed, a prototype based on a virtual **patient** record, is described, and its applications are discussed. TeleMed uses a media-rich graphical **patient** record to allow multiple physicians, possibly located remotely across a wide-area network, to consult on a **patient** record. The latest version of TeleMed supports real-time interactive collaborations between multiple users. ...

...TEXT: Omitted)

Efforts to build a new information infrastructure will enhance and extend the quality of **health** care worldwide. Virtual **patient** records will impact **health** care internationally the way the Internet and the World-Wide Web transformed global communications. At the same time that political and sociological conditions are causing a revolution in **health**

care, advanced technology promise to introduce changes more radical. **Health** care has long lagged far behind other industries in adopting information technologies. More recently, pressures...

... generated an intense interest in how the latest computing technologies can help. While computerized virtual **patient** records [1] could help in many ways, their most significant impact may come from how they change the way we think about international aspects of **health** care. In this article, we describe what we mean by a collaboratory based on virtual **patient** records and the international implications of this new way of managing **health** care data.

Virtual **patient** records are virtual in the sense they provide a view of data possibly configured differently...

... locations, but united together into a common format for viewing at the required time. Virtual **patient** records provide a user with virtual access to data possibly scattered around the world. This...

... the way virtual memory creates larger memory spaces by accessing supplemental storage only when required.

Health care providers and payers have long realized the benefits computer-based **patient** records could provide [2]. Many **health** care practitioners complain that where computer records do exist for their **patients**, the **clinical** data lies tangled in a morass of systems designed primarily for financial claims processing. In...

... did the financial side of our culture computerize its records long ago and leave the **clinical** side undone? Some say it's because **health** care information has much more complexity than financial information. Some say it's because computers require highly structured data, and **health** care data often has inherent ambiguities. Others say economics have higher priority than quality of care. The heart of the matter likely rests near the critical role **health** care plays in human culture.

Many cultural anthropologists define culture as a way of life...

... culture, because it deals with survival. All cultures resist change; it's a natural and **healthy** tendency that helps prevent harmful innovations from destroying a culture. Yet, all cultures do change...

... the more computers represent potentially harmful threats to our lives. Designs for the future of **health** information systems need to preserve as much traditional culture as possible while also providing a...

... means incorporating familiar, and therefore natural, human dynamics such as those involved in physician and **patient** encounters.

With the advent of cyberspace, we now live in a slightly different world. We...

... open to people from all nations, knitting the individual lives of humanity ever closer together. **Health** care is trying to make the most of this new world, too. People in **health** care use real-time video teleconferencing in the practice of medicine and call it "telemedicine" [3]. Some major **hospitals** have even put their **medical** records on the Web [4, 5]. However, these early innovations do not take advantage of some of the more advanced technologies that could provide even better solutions for **health** care while at the same time incorporating the feeling of a traditional live encounter.

Collaboratory...

... the Web provides very important connectivity and access for the publishing of information, it supports **electronic medical records** in a very primitive way compared to other alternatives. The HyperText Transfer Protocol (http), for...

... This provides very good support for interactive connectivity with the data.

We contend the virtual **patient** record should form the basis for a collaborative environment in which multiple physicians, and ultimately the **patient**, should be able to engage in an interactive electronic discussion. This requires the support of...

... transaction processing necessary to support multiple users not only browsing, but also possibly editing a **patient** record simultaneously. As we will explain, the same architecture that enables such realtime consultations can...

... a great enhancement to the typical video consult by providing, in real-time, the full **medical** record of the **patient**. Moreover, by including **patients** and allowing them not only to access their data, but also to post reports and queries, **patients** can become better educated and more empowered to manage their own **health**.

Because of the high degree of complexity of **health** care data and the information intensive aspects of **health** care delivery, these collaborative architectural features are essential for virtual **patient** records. Moreover, this same architecture that supports collaborative consultations in a local region could also... We are collaborating with a number of organizations to help standardize the electronic management of **patient** information, using the most advanced information technologies, to form a seamless international continuum of quality **health** care.

International Implications

Ubiquitous adoption of nonproprietary standards provides the foundation that makes the Internet and the Web possible. By adopting similar international standards for virtual **patient** records, the day may soon come when **electronic medical records** securely cross international borders. In recent years advocates of computer-based **patient** records have set the goal of maintaining longitudinal records that will manage **health** data for a person's entire lifetime-prenatal to post-mortem. **Health** care quality depends on the continuity of care that longitudinal records could provide, but people often cross international borders during their lives, sometimes with the expressed purpose of seeking **health** care outside their own country.

Also, highly intensive monitoring, collecting and mining of **health** data at an international level would better support public policy decisions. Additional information derived from...

...we are collaborating with other organizations to help establish a global network of distributed virtual **patient** records for the surveillance of infectious diseases-the leading cause of death worldwide [6]. Global **health** care surveillance networks could provide important data for international **health** policy decision-making as well as raising the quality of care provided by international relief organizations.

International standards already exist for international electronic exchange of personal financial data. Personal **health** data could move across the networks just as easily based on international standards that provide...

... S. enacted into law a process that will govern the adoption of national standards for **health**-related electronic commerce [7]. The law requires compliance with national standards by the year 2000. This new legislation requires a standard unique **health** identifier for each individual, employer, **health** plan, and **health** care provider for use in the **health** care system. The legislation also calls for other standards to be adopted by the U.S. Department of **Health** and Human Services, and it lays out a timetable for adopting those standards together with...

... some may contend it will be a long time before we see international standards for **health** care. An additional approach to international standardization seeks to base compliance on the collaborative agreements...

...applications happen to be located in the world [8, 9].

The OMG officially established the **health** care domain task force (CORBAMED) to standardize object interfaces for **health** care applications at a global level. A number of other organizations also seek to standardize **health** care information and communication including: CEN (European Committee for Standardization), UN/EDIFACT (United Nations rules for Electronic Data Interchange For Administration, Commerce and Transport), JWG-CDM (**Joint Working Group** for a **Common Data Model**), HL7 (**Health Level 7**), ASTM (American Society for Testing and Materials), CPRI (Computerized **Patient** Record Institute), and others. Some of these standard developing organizations are currently working to develop...

...or Request For Comment (RFC) process.

One of the most basic standardization issues for international **health** care transactions involves how to handle the many different person identifiers existing in disparate **medical** record systems. Even within a single **hospital**, different departments may have separate electronic **records** for the **same** person, each with a different **identification** number for that person. Historically, **health** care providers dealt with this issue by creating a Master **Patient** Index (MPI) that used a limited set of demographic data (for example, name, gender, date of birth, and so on) to help retrieve the disparate elements of a **patient**'s records. Originally, MPIs were manual and were primarily used to link **patient** episodes for continuous care. Maintaining the continuity of **patient** care is still the most important role of automated MPIs. However, existing MPI software applications...

... to organize a series of workshops aimed directly at this issue. Viability of a virtual **patient** record collaboratory will depend heavily on the ability to quickly and securely identify a **patient** and his or her respective **health** care providers and payers. Different indexing schemes used in different nations, as well as privacy MPI systems. The **Health Level 7** standards developing organization has now started a special interest group focusing on MPI Mediator issues.

A MPI can be summarized as a secured directory of uniquely identifiable **patients** together with pointers to locations where the respective **medically** related information about each **patient** resides. A MPI does not record any **clinical** data; it merely points to where **clinical** data is located, whether in electronic form or possibly on paper. By relying on proprietary...

...grant access to pointers that provide authorized and secure access paths to records from disparate **health** care organizations, even when distributed at an international level.

Pointers to sites of care have the potential for revealing confidential information about a **patient**; a pointer could reveal a **patient** received care at an AIDS clinic, an abortion clinic, or in prison. Therefore, privacy advocates contend that standards for authorization and access controls must allow the **patients** to decide to whom they grant access and which records, if any, to make available...

... 9]. The MPI Mediator approach has the potential to facilitate such authorizations for access to **health** records.

After helping to organize the series of workshops seeking a solution to the MPI...

... international level-we worked closely with CORBAMED to help generate a RFP requesting proposals for **patient** identification services. We are currently collaborating with various MPI vendors to encourage them to produce...

... the final specific on the OMG recommends could come the key to mediating international exchanges of **health** - related information.

TeleMed: A Prototype Based on a Virtual **Patient** Record

Working in collaboration with physicians at the National Jewish **Medical** and Research Center (NJC) in Denver, we developed TeleMed, an early prototype for a virtual **patient** record. TeleMed uses a media-rich graphical **patient** record [9] to allow multiple physicians, possibly located remotely across a wide-area network, to consult on a **patient** record. Consultations can take place interactively in realtime, or offline using textual or audio annotations combined with graphical markers in the record.

The original prototype demonstrated the feasibility of virtual **patient** records, but it did not reach the stage of full functionality with actual **patient** records on a day-to-day basis. Los Alamos National Laboratory plans to deploy a more functional version of TeleMed in some of the small clinics and **hospitals** of northern New Mexico. The Northern New Mexico Rural Telemedicine Project (NNM RTP) led by the...

...for a collaboratory based on the TeleMed prototype.

TeleMed dynamically assembles a chronologically oriented graphical **patient** record using data gathered from several different remote locations. It assigns icons and other graphical...

... Object Request Brokers (ORBs) then mediate the interactive access linking the icons in the graphical **patient** record to the distributed databases that provide the persistent object storage for the multimedia data...

... between multiple users. Multiple physicians at remote locations can simultaneously view, edit, and annotate the **patient** data. Furthermore, each physician can see the data another physician has entered as well as...

... enables the support of full transaction processing necessary to support multiple users interacting with the **patient** record simultaneously. Such a capability enables the same architecture to be used for real-time...

... different database. Once a database has been selected, a CORBA-based transaction vends the relevant **patient** record objects from the selected site to the requesting client. Textual data from the retrieved **patient** objects is then used to populate a list of **patient** names.

When the user selects one of the names in the list, that **patient**'s Graphical **Patient** Record (GPR) window appears. TeleMed's chronological GPR is dynamically assembled by CORBA-based transactions that **merge** all of the relevant **data** from **multiple** databases to form the seamless view presented to the user. The fact the user selected...

... as a guide to help determine where to produce full transverse slice images of the **patient**. In this interface, the scout is now used as a guide to help the user...

... particular interface. Suppose a general practice physician wanted to consult with a specialist about this **patient** by using TeleMed. And let's say the physician wants to discuss the particular slice...

...display window depending on the type of annotation.

Conclusion

An international collaboratory based on virtual **patient** records can eventually include a wide assortment of **health** care applications all designed to allow **health** practitioners and their **patients** to work together more naturally and more effectively and, thereby, deliver a higher quality care...

... to invest now during time of transition to develop new information infrastructure for health care.

More information about TeleMed can be found at
<http://www.acl.lanl.gov/TeleMed>

Reference:
REFERENCES

Reference:

1. Forslund, D. W. et al. Experiences with a distributed virtual patient record system. In Proceedings of the 1996 American Medical Informatics Association Annual Fall Symposium (Washington, D.C., Oct. 26-30). J. J. Cimino, Ed. American Medical Informatics Assoc., Hanley & Belfus, 1996, pp. 483-487.
2. Dick, R. S., and Steen E. B., Eds. The Computer-Based Patient Record, An Essential Technology for Health Care. Institute of Medicine, National Academy Press, Washington, D.C., 1991.
3. Field, M. J., Ed. Telemedicine: A Guide to Assessing Telecommunications in Health Care, Institute of Medicine. National Academy Press, Washington, D.C., 1996.
4. Kohane, I. A., et al. Sharing electronic medical records across multiple heterogeneous and competing institutions. In Proceedings of the 1996 American Medical Informatics Association Annual Fall Symposium (Washington, D.C., Oct. 26-30). J. J. Cimino, Ed. American Medical Informatics Assoc., Hanley & Belfus, 1996, pp. 608-612.
5. McDonald, C. J., et al. The Regenstrief medical record system (RMRS): Physician use for input and output and Web browser based computing. In Proceedings of the 1996 American Medical Informatics Association Annual Fall Symposium (Washington, D.C., Oct. 26-30). J. J. Cimino, Ed. American Medical Informatics Assoc., Hanley & Belfus, 1996, p. 989.
6. Center for Disease Control and Prevention. National...
...Request Broker Architecture. Wiley-QED, 1991.
10. Donaldson, M. S., and Lohr, K. N., Eds. Health Data in the Information Age: Use, Disclosure, and Privacy. Institute of Medicine, National Academy Press...

DESCRIPTORS: Medical records...

... Health care
?

22/3,K/2 (Item 2 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

02223566 80529043

The utility and selection of an HRIS

Targowski, Andrew S; Deshpande, Satish P
Advances in Competitiveness Research v9n1 PP: 42-56 2001
ISSN: 1077-0097 JRNL CODE: AICR
WORD COUNT: 4160

...TEXT: management skills (Roberts, 1999).

Privacy Issues

Employers have the obligation to protect employees' personal records, **medical** records, and the results of employment related tests (Myers, 1988). Major privacy concerns related to...

... information can be stored in the system and who can access, change, or modify the **database** (Noe, Hollenbeck, Gerhart, and Wright, 1994). It is suggested that a company should only collect...

... to those who had physical access to it. But today an HRIS uses a relational **database** to store and categorize different data files (Hubbard, Forcht, and Thomas, 1998). Relational **databases** not only allow the user to access and **merge information** via a **common** element like name or employee **identification** number, it also allows this information to be accessed from different locations within the company...

22/3,K/7 (Item 7 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

01839820 04-90811

Streamline the registration process with EMPI

Roget, Melisa
Health Management Technology v20n5 PP: 34-37 Jun 1999
ISSN: 1074-4770 JRNL CODE: CIH
WORD COUNT: 1637

...TEXT: the interface engine and sends information back to the registration system.

(Photograph Omitted)

If the **patient** does not appear on the list, the user enters information into the EMerge **database**. EMerge assigns a **medical** record number. The registration process is the only place in the enterprise where a **medical** record number can be assigned.

Inbound TSS API calls are made to EMerge from workstations...

... into EMerge from the existing Eclipsys 7000 registration system. Out of more than one million **patient** records, over 75,000 possible duplicate records were reviewed between January and September. More than 35,000 of these **records** were **identified** as **duplicates** and were **merged** -- reducing the EMerge **database** to more than 1.6 million records. As the project grew to include multiple implementations, UAB decided additional resources were needed. The **healthcare** consulting firm, Insource Management Group, was enlisted to provide project management.

Phase 2: As of...

22/3,K/10 (Item 10 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

01624131 02-75120

How to find who

Cupito, Mary Carmen

Health Management Technology v19n6 PP: 32-33+ May 1998

ISSN: 1074-4770 JRNL CODE: CIH

WORD COUNT: 1787

...TEXT: enterprise.

Yet, the enterprisewide indexing application itself is often expected to do a lot: receive **patient** or member information from many systems, assign unique identifiers to each individual, **merge records** pertaining to the **same** person, update its **database**, **remove duplicate records**, correlate a **patient**'s records throughout the enterprise, keep the source system's identifier or pointers to it...automatically merges them.

"Physically," the records are not merged within the global master person index **database**, so that if it is later determined that these records are not for the same **patient**, the records can be 'unmerged' and restored to the original state," Frede noted. "We feel...

...s own network that person's data is stored. "We take the data into our **database** and we link it to a single **patient** across all systems," Frede says. "We're not replacing any existing system. We consider ourselves...

22/3,K/22 (Item 22 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

00725639 93-74860

Cleanup and Deduplication of an International Bibliographic Database

Toney, Stephen R.

Information Technology & Libraries v11n1 PP: 19-28 Mar 1992

ISSN: 0730-9295 JRNL CODE: JLA

WORD COUNT: 5235

...TEXT: s own.

REFERENCES

1. Martha E. Williams and Keith D. MacLaury, Automatic Merging of Monographic **Data Bases Identification of Duplicate Records in Multiple Files** : The IUCS Scheme," Journal of Library Automation 12, no.2:156-68 June 1979). 2. June E. Slach, "Detection and Elimination of Duplicates from Multidatabase Searches," Bulletin of the **Medical Library Association** 73, no.3:235-37 (July 1985). 3. C. A. Giles and others, A Computerized Scheme for **Duplicate Checking of Bibliographic Data Bases** (NTIS CRNL/CSD-5, Aug. 1976). 4. C. A. Giles and others, An Experiment in...
... CONF-76-1011-2. n.d.). 5. Thomas B. Hickey and David J. Rypka, "Automatic **Detection of Duplicate Monographic Records**," Journal of Library Automation 12, no.2:125-42 June 1979). 6. James L. Dolby...

...Retrieval," Journal of Library Automation 1, no.4:250-60 (Dec. 1968). 8. Pankaj Goyal, " **Duplicate Record Identification in Bibliographic Databases**," Information Systems 12, no.3:23942 (1987). 9. E. J. Yannakoudakis and others, "Matching of Citations Between Non-Standardized **Databases**," Journal of the American Society for Information Science 41, no.8:599-610 (1990). 10...

22/3,K/43 (Item 1 from file: 613)

00383761 20000727SFTH096 (USE FORMAT 7 FOR FULLTEXT)

The American Association for Artificial Intelligence (Aaai) Awards The Best in Practical Innovations of Intelligent Systems Applications

PR Newswire

Thursday, July 27, 2000 14:00 EDT

JOURNAL CODE: PR LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT

DOCUMENT TYPE: NEWSWIRE

WORD COUNT: 687

...adopted into other

hospitals across Hong Kong.

Title: "LifeCode: A Natural Language Processing System for
Medical

Coding and **Data Mining** "

Submitted by: A-Life **Medical**, Inc. (San Diego, California. USA)

Summary: The LifeCode Expert System (patent pending) is a

Natural

Language Processing (NLP) based engine that extracts and provides useful data for the Emergency Medicine

clinical

specialty. Its outputs provide an integrated

illustration

of the **patients' medical** condition(s), course of

treatment

and disposition while intelligently **deleting**

redundant

information, combining complementary **information**

to create

the integrated picture, and assessing the **medical**

risks and

service. LifeCode also has a "self-awareness" feature

that

recognizes its own limits...

?

?t s25/3,k/8,11,15,35,46,67,73,84

25/3,K/8 (Item 8 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

01148712 97-98106

UnInstaller makes strong entry into Win95 market

Peschel, Joe

InfoWorld v18n4 PP: 95 Jan 22, 1996

ISSN: 0199-6649 JRNL CODE: IFW

WORD COUNT: 615

...TEXT: Cleanup: Tracks and deletes files you no longer need.

* Setup Monitor: Tracks new program installations.

* **Duplicate File Finder** : Finds duplicates by name/ date /size combinations , size filters.

* Disk Data: Finds folders and files; also gives a report of disk space...

25/3,K/11 (Item 11 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

00922670 95-72062

To merge and not to merge: Israel's union list of monographs in the context of merging algorithms

Lazinger, Susan S

Information Technology & Libraries v13n3 PP: 213-219 Sep 1994

ISSN: 0730-9295 JRNL CODE: JLA

WORD COUNT: 5495

...TEXT: to bring together records derived from a standard LC cataloging in a minimum of processing time . A **second** phase processed **records** that didn't **match** on the minimum set of data elements through the remainder of the algorithm, comparing the...

25/3,K/15 (Item 15 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

00725639 93-74860

Cleanup and Deduplication of an International Bibliographic Database

Toney, Stephen R.

Information Technology & Libraries v11n1 PP: 19-28 Mar 1992

ISSN: 0730-9295 JRNL CODE: JLA

WORD COUNT: 5235

...TEXT: trail of great value when programming errors or additional cleanup requirements were noted. Furthermore, each **duplicate record merged** was noted in its master record by number and **date** . Both these ideas are recommended.

Finally, never underestimate the time it takes for a project...

25/3,K/35 (Item 6 from file: 275)
DIALOG(R)File 275:Gale Group Computer DB(TM)
(c) 2004 The Gale Group. All rts. reserv.

02162894 SUPPLIER NUMBER: 20502978 (USE FORMAT 7 OR 9 FOR FULL TEXT)

Don't Waste Disk Space. (Dupeless utility for Borland's Delphi) (Product Information)

Rubenking, Neil J.
PC Magazine, v17, n7, p259(1)
April 7, 1998
ISSN: 0888-8507 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 4126 LINE COUNT: 00311

... file (same name and size);
* identical copies of the same file (same name, size, and **contents**);
* unexpected **identical files** (different names but **same size and contents**).

The **Comparison** options box is where you make your selections.

Either filename or file size must be...filenames; wildcards are not allowed. The list of extensions starts off containing . and .TMP, both **commonly** used for temporary files . (It's quite **common** to find temporary files whose **contents** are **identical** to those of the corresponding permanent files.) Again, wildcards are not allowed.

If you wish...

...that were used in comparing files.

You'll probably be amazed at the number of **duplicate files** Dupeless **finds** the first **time** you use it. And even though the prices of immense hard disks continue to drop...on a variety of definitions, and letting it do so in something less than geologic **time** . Speedily **matching up files with identical contents** was a separate challenge, solved in part by having Dupeless defer the actual file-contents...

...all of the files in the temporary list as a group of duplicates--unless the **Contents** option has been selected.

Finding Identical Files

If the **Contents** option is selected, Dupeless must go through another step before declaring a set of files...

25/3,K/46 (Item 17 from file: 275)
DIALOG(R)File 275:Gale Group Computer DB(TM)
(c) 2004 The Gale Group. All rts. reserv.

01713196 SUPPLIER NUMBER: 15040443 (USE FORMAT 7 OR 9 FOR FULL TEXT)
The 29-cent stomp. (SoftSpoken's SoftSpoken Mailer for Windows 2.0 mailing list and data base manager) (Software Review) (Evaluation)
Hollander, Geoffrey
Home Office Computing, v12, n3, p32(2)
March, 1994
DOCUMENT TYPE: Evaluation ISSN: 0899-7373 LANGUAGE: ENGLISH
RECORD TYPE: FULLTEXT; ABSTRACT
WORD COUNT: 741 LINE COUNT: 00057

... ready to go. The program seamlessly accesses, indexes, and stores file configurations, mailing output, and **dupe (duplicate record) information** .

Sorting, **combining** , and packaging (called presorting) mail for rapid disbursement is a **time** -consuming task--no matter who does it. But the USPS has long offered substantial postage...

...are simultaneously checked, but the program will allow you to assign any number of field **combinations** . Searches will locate exact **duplicates (records whose contents match down to the letter)** as well as the more elusive near-dupes (records that, due...

25/3,K/59 (Item 30 from file: 275)
DIALOG(R)File 275:Gale Group Computer DB(TM)
(c) 2004 The Gale Group. All rts. reserv.

01528449 SUPPLIER NUMBER: 12351740 (USE FORMAT 7 OR 9 FOR FULL TEXT)
DupLocator digs up duplicate disk files. (Software Review) (Midnight Software Inc. DupLocator utility for disk organization) (includes related summary article) (Evaluation)

Shapiro, Neil

MacWEEK, v6, n26, p84(3)

July 13, 1992

DOCUMENT TYPE: Evaluation

ISSN: 0892-8118

LANGUAGE: ENGLISH

RECORD TYPE: FULLTEXT; ABSTRACT

WORD COUNT: 1020 LINE COUNT: 00077

...ABSTRACT: a window that lists only duplicated files; DupLocator examines only file names and not their **contents**, forcing the user to **match** information to weed out true **duplicates**. Each **file**'s disk path is displayed in a large window below the main window. DupLocator has...

... Quadra 700.

Seeing doubles. The View Duplicate Files command creates a window that lists only **duplicated files**. **Finding** the true **duplicates** then becomes the main task. DupLocator doesn't examine the **contents** of a file -- duplicates are based solely upon the file name.

The main window displays...

25/3,K/67 (Item 38 from file: 275)

DIALOG(R)File 275:Gale Group Computer DB(TM)

(c) 2004 The Gale Group. All rts. reserv.

01436240 SUPPLIER NUMBER: 10864238 (USE FORMAT 7 OR 9 FOR FULL TEXT)

Info Select 2.0 adds neural search, LAN option, info base sorting, and

quick envelope printing. (Micro Logic Corp.) (product announcement)

Computer Shopper, v11, n7, p508H(1)

July, 1991

DOCUMENT TYPE: product announcement

ISSN: 0886-0556

LANGUAGE:

ENGLISH RECORD TYPE: FULLTEXT

WORD COUNT: 279 LINE COUNT: 00022

... key word. Also, an ascending or descending Windows Sort can sort by string, number, or **date**.

Un-Dupe is a "**remove duplicates**" capability that looks for **duplicate information** in a **merged** database. Envelopes and Shipping Label Printing automatically recognizes the address portion of a window and...

25/3,K/73 (Item 44 from file: 275)

DIALOG(R)File 275:Gale Group Computer DB(TM)

(c) 2004 The Gale Group. All rts. reserv.

01261299 SUPPLIER NUMBER: 07106501 (USE FORMAT 7 OR 9 FOR FULL TEXT)

The data chase. (Macuser Labs evaluates seven relation databases) (Software

Review) (one of four articles in a special section: The Information

Machine) (includes related article on Hypercard as a visual database;

other related articles discuss databases with multiuser capabilities and

A-UX's impact on DBMS) (evaluation)

Young, Jeffrey; Warren, Carl

MacUser, v4, n12, p159(22)

Dec, 1988

DOCUMENT TYPE: evaluation

ISSN: 0884-0997

LANGUAGE: ENGLISH

RECORD TYPE: FULLTEXT; ABSTRACT

WORD COUNT: 8099 LINE COUNT: 00626

... of the products tested allow you to select a record or file and then to **delete** it. We **deleted multiple records** by **removing** all customers older than a fixed entry **date**.

Infrequently, you may find that you need to make changes in your data files that...

25/3,K/84 (Item 6 from file: 636)

DIALOG(R)File 636:Gale Group Newsletter DB(TM)

(c) 2004 The Gale Group. All rts. reserv.

02444688 Supplier Number: 44880082 (USE FORMAT 7 FOR FULLTEXT)

EBSCO PUBLISHING RELEASES MEDLINE/EBSCO CD-ROM VERSION 4.0

CD Computing News, v8, n8, pN/A

August, 1994

Language: English Record Type: Fulltext

Document Type: Newsletter; Trade

Word Count: 470

... either to the hard drive or a floppy diskette. Results from several searches can be **combined** in a single **file** (**duplicate records** are automatically **removed**) and the results can be printed at a later **date** . The SDI feature has been enhanced to allow batch processing of complex queries which include...

?

?ds.

Set	Items	Description
S1	1559170	DUPLICAT? OR DUPE? ? OR COMMON? OR SAME? OR REDUNDAN? OR R- EPLICA? OR REPEAT? OR IDENTICAL? OR DOUBLE? OR ANALOGOUS? OR - SECOND OR PARALLEL? OR MULTIPLE?
S2	721154	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?.
S3	723805	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION OR CITATION?
S4	1427234	REMOV? OR DELETE? OR DELETING? OR ERASE? ? OR ERASING? OR - DEDUPE? ? OR CHECK? OR FIND???? OR IDENTIF? OR DETECT? OR C- OMPAR???? OR MATCH? OR PREVENT?
S5	153441	S1(3N)S3
S6	21748	S5(5N)S4
S7	1255200	MERGE? OR MERGING? OR CONSOLIDAT? OR COMBIN? OR JOIN??? OR GROUP? OR COMPRESS? OR COMPOSITE?
S8	12904	S7(5N)S5
S9	3483	S6 AND S8
S10	300852	MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL? OR CLINICAL?
S11	510	S10 AND S9
S12	36	S11 AND IC=(G06F-007/00 OR G06F-017/30)
S13	52	S9 AND S10/TI,AB
S14	25	S13 AND IC=G06F?
S15	34	S12 NOT S14

?show files

File 348:EUROPEAN PATENTS 1978-2004/Mar W03

(c) 2004 European Patent Office

File 349:PCT FULLTEXT 1979-2002/UB=20040325,UT=20040318

(c) 2004 WIPO/Univentio

?

14/3,K/1 (Item 1 from file: 348)
DIALOG(R) File 348:EUROPEAN PATENTS
(c) 2004 European Patent Office. All rts. reserv.

01535460

A method for generating a user interface

Verfahren zur Erzeugung einer Benutzerschnittstelle

Procede pour la creation d'une interface utilisateur

PATENT ASSIGNEE:

SCA Hygiene Products AB, (2569621), Backstensgatan 5, Molndal, 405 03
Goteborg, (SE), (Applicant designated States: all)

INVENTOR:

Kling, Asa, Alegardsgatan 265, 431 50 Molndal, (DK)
Svensson, Bjorn, Klarinettvagen 4, 435 41 Molnlycke, (SE)

LEGAL REPRESENTATIVE:

Plougmann & Vingtoft A/S (101177), Sundkrogsade 9, P.O. Box 831, 2100
Copenhagen, (DK)

PATENT (CC, No, Kind, Date): EP 1280085 A1 030129 (Basic)

APPLICATION (CC, No, Date): EP 2002078062 020725;

PRIORITY (CC, No, Date): US 911733 010725

DESIGNATED STATES: AT; BE; BG; CH; CY; CZ; DE; DK; EE; ES; FI; FR; GB; GR;
IE; IT; LI; LU; MC; NL; PT; SE; SK; TR

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI

INTERNATIONAL PATENT CLASS: **G06F-017/60**

ABSTRACT WORD COUNT: 123

NOTE:

Figure number on first page: 4

LANGUAGE (Publication,Procedural,Application): English; English; English

FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS A	(English)	200305	849
SPEC A	(English)	200305	6135
Total word count - document A			6984
Total word count - document B			0
Total word count - documents A + B			6984

INTERNATIONAL PATENT CLASS: **G06F-017/60**

...ABSTRACT A1

A method for generating a user interface for displaying data, such as **healthcare** product consumption and cost information, comprises allocating a plurality of institutions, e.g. **hospitals** or nursing homes, into groups. Groups may be altered, and time periods and currencies may...

...SPECIFICATION group of the plurality of groups of institutions, identification data of institutions contained in that **group** ,
- creating a **second** set of **data** in response to a second user input, the **second** set of **data** representing at least an **identification** of the institution group, the time period, and the currency type selected by the user...group of the plurality of groups of institutions, identification data of institutions contained in that **group** ,
- create a **second** set of **data** in response to a second user input, the **second** set of **data** representing at least an **identification** of the institution group, the time period, and the currency type selected by the user...

...CLAIMS group of the plurality of groups of institutions, identification data of institutions contained in that **group** ,
- creating a **second** set of **data** in response to a second user input, the **second** set of **data** representing at least an **identification** of the institution group, the time period, and the currency type selected by the user...

...group of the plurality of groups of institutions, identification data of institutions contained in that group ,
- create a second set of data in response to a second user input, the second set of data representing at least an identification of the institution group, the time period, and the currency type selected by the user...

14/3,K/2 (Item 2 from file: 348)
DIALOG(R) File 348:EUROPEAN PATENTS
(c) 2004 European Patent Office. All rts. reserv.

01281461

Automated collection and analysis patient care system and method for ordering and prioritizing multiple health disorders to identify an index disorder

Automatisiertes Sammlungs und Analyse- Patienten ,Pflegesystem und Verfahren zum Ordnen und setzen von Prioritaten fur ein Vielfach von Krankheiten zum Identifizieren einer Indexkrankheit

Systeme de gestion de soins aux patients avec collection et analyse automatique et methode d'ordonnancement et d'etablissement de priorites dans un multiple de maladies pour identifier une maladie index

PATENT ASSIGNEE:

Cardiac Intelligence Corporation, (3179130), 2518 Constance Drive West, Seattle, Washington 98199-3017, (US), (Applicant designated States: all)

INVENTOR:

Bardy, Gust H., 2518 Constance Drive West, Seattle, Washington 98199-3017, (US)

LEGAL REPRESENTATIVE:

Hanna, Peter William Derek et al (72343), Hanna, Moore & Curley, 11 Mespil Road,, Dublin 4, (IE)

PATENT (CC, No, Kind, Date): EP 1102200 A2 010523 (Basic)
EP 1102200 A3 020605

APPLICATION (CC, No, Date): EP 2000650198 001116;

PRIORITY (CC, No, Date): US 441405 991116

DESIGNATED STATES: AT; BE; CH; CY; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI; LU; MC; NL; PT; SE; TR

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI

INTERNATIONAL PATENT CLASS: G06F-019/00

ABSTRACT WORD COUNT: 150

NOTE:

Figure number on first page: 1

LANGUAGE (Publication,Procedural,Application): English; English; English

FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS A	(English)	200121	1664
SPEC A	(English)	200121	8454
Total word count - document A			10118
Total word count - document B			0
Total word count - documents A + B			10118

Automated collection and analysis patient care system and method for ordering and prioritizing multiple health disorders to identify an index disorder

Automatisiertes Sammlungs und Analyse- Patienten ,Pflegesystem und Verfahren zum Ordnen und setzen von Prioritaten fur ein Vielfach von Krankheiten zum...

Systeme de gestion de soins aux patients avec collection et analyse automatique et methode d'ordonnancement et d'etablissement de priorites dans...

INTERNATIONAL PATENT CLASS: G06F-019/00

...ABSTRACT A2

An automated collection and analysis patient care system and method for ordering and prioritizing multiple health disorders to identify an

index disorder is described. A plurality of monitoring sets are retrieved from a database. Each of the monitoring sets include stored measures relating to **patient** information recorded and derived on a substantially continuous basis. A **patient** status change is determined by comparing at least one stored measure from each of the...

...least one other stored measure with both stored measures relating to the same type of **patient** information. Each **patient** status change is ordered in temporal sequence from least recent to most recent. A plurality of **health** disorder candidates categorized by quantifiable physiological measures of pathophysiologies indicative of each respective **health** disorder are evaluated and the **health** disorder candidate with the pathophysiology most closely matching those **patient** status changes which occurred least recently is identified.

...SPECIFICATION or improvement of cardiovascular disease. Additionally, the Nappholz device is unable to automatically self-reference **multiple data** points in time and cannot **detect** disease regression. Also, the Nappholz device must be implanted and cannot function as an external...as Patient 1, serial peer measures, such as peer measure X₀ through X_N), represent the **same** type of patient **information** monitored over time. The **combined** peer measures for all patients can be categorized into a health disorder- or disease-matched...

...CLAIMS each patient status change for each stored measure to the stickiness indicator relating to the **same** type of patient **information** as the stored measure being **compared**; and the analysis module (207) determining a revised program of patient diagnosis or treatment responsive...each patient status change for each stored measure to the stickiness indicator relating to the **same** type of patient **information** as the stored measure being **compared**; and determining a revised program of patient diagnosis or treatment responsive to each patient status...

14/3,K/10 (Item 1 from file: 349)

DIALOG(R)File 349:PCT FULLTEXT

(c) 2004 WIPO/Univentio. All rts. reserv.

01095251 **Image available**

METHODS AND SYSTEMS FOR MANAGING DISTRIBUTED DIGITAL MEDICAL DATA AND ACCESS THERETO

PROCEDES ET SYSTEMES DE GESTION DE DONNEES MEDICALES NUMERIQUES DISTRIBUEES ET ACCES A CES DONNEES

Patent Applicant/Assignee:

HX TECHNOLOGIES INC, 9 Hawthorne Avenue, Morris Plains, NJ 07950, US, US
(Residence), US (Nationality)

Inventor(s):

MENSCHIK Elliot, 9 Hawthorne Avenue, Morris Plains, NJ 07950, US,
CORIO Christopher, 2801 Western Avenue #623, Seattle, WA 98121, US,
DAVIS Wayne, 600 Wesley Avenue, Ocean City, NJ 08226, US,
DIDIZIAN Haig, 400 Old Gulph Road, Penn Valley, PA 19072, US,

Legal Representative:

BRANDT Jeffrey (agent), Axiom Legal Solutions c/o PortfolioIP, P.O. Box 52050, Minneapolis, MN 55402, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200417164 A2 20040226 (WO 0417164)

Application: WO 2003US24908 20030807 (PCT/WO US03024908)

Priority Application: US 2002222056 20020816; US 2002222720 20020816

Designated States: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU

CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP

KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NI NO NZ OM PG PH PL

PT RO RU SC SD SE SG SK SL SY TJ TM TN TR TT TZ UA UG UZ VC VN YU ZA ZM

ZW

(EP) AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL PT RO SE

SI SK TR

(OA) BF BJ CF CG CI CM GN GQ GW ML MR NE SN TD TG
(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZM ZW
(EA) AM AZ BY KG KZ MD RU TJ TM

Publication Language: English

Filing Language: English

Fulltext Word Count: 25573

METHODS AND SYSTEMS FOR MANAGING DISTRIBUTED DIGITAL MEDICAL DATA AND ACCESS THERETO

PROCEDES ET SYSTEMES DE GESTION DE DONNEES MEDICALES NUMERIQUES DISTRIBUEES ET ACCES A CES DONNEES

Main International Patent Class: G06F

Fulltext Availability:

Detailed Description

Claims

English Abstract

A network (100/160/180) for mediating the peer-to-peer transfer of digital **patient medical** data includes a plurality of distributed agents (180A, 180B, 180E, 180F, 180G, 180H) each associated with a **health** care provider and connected to a central system (180C). Periodically the agents (180A, 180B, 180E, 180F, 180G, 180H) collect local information relating to **patient medical** files and/or data streams, for example diagnostic images and associated reports, and process that...

...where they are parsed and the attributes are stored on the central system (180C) in **patient** records with **records** from the **same patient grouped** together whenever possible. Registered users can search the central system (180C), even in the absence of a unique identifier, to identify **patient** records pointing to the remote **patient medical** files. Upon finding a **patient medical** file, the invention provides a streamlined process for communicating access authorization from the **patient** to the **hospital** or facility storing the **medical** files. Once **patient** authorization is received, secure processes are provided for transferring the data in its entirety to...

French Abstract

...reseau (100/160/180) destine a induire le transfert d'egal a egal de donnees **medicales** numeriques d'un **patient**, ces donnees comprenant une pluralite d'agents distribues (180A, 180B, 180E, 180F, 180G, 180H), chacun...

...180A, 180B, 180E, 180F, 180G, 180H) collectent les informations locales relatives aux fichiers medicaux des **patients** et/ou aux trains de donnees, par exemple, des images de diagnostic et des rapports...

...analyses et les attributs sont stockes sur le systeme central (180C) dans des registres de **patients**, avec des enregistrements du meme **patient** regroupes chaque fois que cela est possible. Les utilisateurs enregistres peuvent rechercher le systeme central (180C), meme en l'absence d'un identificateur unique, afin d'identifier des enregistrements de **patients** designant les fichiers medicaux a distance des **patients**. Pour rechercher un fichier **medical** d'un **patient**, on peut utiliser un processus en continu de l'invention permettant de communiquer l'autorisation d'accès du **patient** a l'hospital ou a l'installation stockant les fichiers medicaux. A la reception de l'autorisation du **patient**, des processus de securite sont mis en place pour transferer les donnees dans leur totalite...

Detailed Description

... not propose an actual solution, but merely suggests that it should be possible to use **multiple** pieces of patient **information** to **match** patients.

...containing attributes relating to a remotely stored patient medical file; extracting the attributes from the **second metadata file** ;
comparing , by the computer, the attributes from the second metadata file to the attributes stored in

Claim

... containing attributes relating to a remotely stored patient medical file;
extracting the attributes from the **second metadata file** ;
comparing , by the computer, the attributes from the second metadata file to the
attributes stored in...

...the
database entry.

47

. A method in accordance with claim 26 wherein the step of **comparing** includes the steps of parsing the **second metadata file** to extract the attributes contained therein; and matching at least one attribute extracted from the...

...determining if the second metadata file relates to the database entry includes determining if the **second metadata file** is a **match** , a non-**match** or indeterminate with respect to the database entry.

29 A method in accordance with claim 28 and further comprising the steps of: if the **second metadata file matches** the database entry, updating the database entry; and
if the **second metadata file** is a non- **match** with the database entry, creating a new database entry based on the attributes in the...

...containing attributes relating to a remotely stored patient medical file;
extracting the attributes from the **second metadata file** ;
comparing the attributes from the **second metadata file** to the attributes stored
in the database entry; and
determining if the second metadata file...

...determining if the second metadata, file relates to the database entry includes determining if the **second**
49@'
metadata **file** is a **match** , a non- **match** or indeterminate with respect to the database entry.

36 A system in accordance with claim 35 and further comprising the steps of: if the **second metadata file matches** the database entry, updating the database entry; and
if the **second metadata file** is a non- **match** with the database entry, creating a new database entry based on the attributes in the...

...containing attributes relating to a remotely stored patient medical file;
extracting the attributes from the **second metadata file** ;
comparing the attributes from the **second metadata file** to the attributes stored
in the database entry; and
1 0 determining if the second...

...relating to a
remotely stored patient medical file;
means for extracting the attributes from the **second metadata file** ;
means for **comparing** the attributes from the **second metadata file** to the
attributes stored in the database entry; and

means for determining, the computer...

...containing attributes relating to a remotely
stored patient medical file;
extracting the attributes from the **second metadata file** ;
I O **comparing** , by the computer, the attributes from the second metadata
file to the
attributes stored in...

14/3,K/15 (Item 6 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2004 WIPO/Univentio. All rts. reserv.

01008726 **Image available**

A HEALTHCARE **SYSTEM AND USER INTERFACE FOR CONSOLIDATING PATIENT
RELATED INFORMATION FROM DIFFERENT SOURCES**

**SYSTEME DE SANTE ET INTERFACE UTILISATEUR PERMETTANT DE CONSOLIDER DES
INFORMATIONS DE PATIENT PROVENANT DE SOURCES DIFFERENTES**

Patent Applicant/Assignee:

SIEMENS MEDICAL SOLUTIONS HEALTH SERVICES CORPORATION, 51 Valley Stream
Parkway, Malvern, PA 19355, US, US (Residence), US (Nationality)

Inventor(s):

MAUGHAN Rex Wendell, 5926 South Normandy Oaks Circle, Murray, UT 84123,
US,

JACOBSEN Jeffry Brent, 1465 East 1425 North, Logan, UT 84321, US,

Legal Representative:

BURKE Alexander J (et al) (agent), Siemens Corporation, Intellectual
Property Dept., 186 Wood Avenue South, Iselin, NJ 08830, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200338731 A2-A3 20030508 (WO 0338731)

Application: WO 2002US34884 20021030 (PCT/WO US02034884)

Priority Application: US 2001335976 20011031; US 2002282644 20021029

Designated States: CA JP

(EP) AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LU MC NL PT SE SK TR

Publication Language: English

Filing Language: English

Fulltext Word Count: 10992

A HEALTHCARE **SYSTEM AND USER INTERFACE FOR CONSOLIDATING PATIENT
RELATED INFORMATION FROM DIFFERENT SOURCES**

**SYSTEME DE SANTE ET INTERFACE UTILISATEUR PERMETTANT DE CONSOLIDER DES
INFORMATIONS DE PATIENT PROVENANT DE SOURCES DIFFERENTES**

Main International Patent Class: G06F-019/00

Fulltext Availability:

Detailed Description

Claims

English Abstract

In a **healthcare** system, a method **consolidates patient** related
information from **multiple** , different sources by performing the
following steps. **Patient** identification data identifying a particular
patient is received. Messages are generated for communication to a
corresponding information sources. The generated messages incorporate the
particular **patient** identification data and a request for information
concerning the particular **patient** . The messages are communicated to the
corresponding information sources. Response messages are received
containing requested information concerning the particular **patient** from
the information sources. The requested information is sorted and merged
from the response messages to provide data representative of consolidated
patient related information.

French Abstract

...invention concerne un systeme de sante utilisant un procede destine a
consolider des informations de **patient** en provenance de sources
differentes multiples. Ce procede consiste a recevoir des donnees
d'identification de **patient** permettant d'identifier un **patient**

particulier, a produire les messages en vue de leur communication a des sources d'informations correspondantes, ces messages comprenant les donnees d'identification du **patient** particulier ainsi qu'une demande d'informations concernant ce **patient** particulier, a communiquer ces messages aux sources d'informations correspondantes, a recevoir des messages de reponse contenant les informations demandees sur le **patient** particulier en provenance des sources d'informations, puis a trier et fusionner les informations demandees...

...de ces messages de reponse en vue d'obtenir des donnees representatives d'informations de **patient** consolidees.

Detailed Description

... services by many healthcare

generate a correct match for over 90% of the stored patient **information**, generate a **duplicate match** for 5% to 10% of the stored patient information, and generate less than 5% of...of the large number of variables, some related to human interaction.

Still other present systems **combine** the patient **information** from the **multiple**, different healthcare sources into a single view. The single view makes is very difficult to...a healthcare information system.

3

Summary of the Invention

In a healthcare system, a method **consolidates** patient related **information** from **multiple**, different sources by performing the following steps. Patient identification data identifying a particular patient is...means to combine the various replies, and may include, without limitation, sorting, manipulating, formatting, purging **duplicate data**, **merging**,

18

and organizing the information. Hence, the server device 14 combines the received information in...

Claim

1. A method for use in a healthcare system for **consolidating** patient related **information** from **multiple** different sources, comprising the steps of: receiving patient identification data identifying a particular patient; generating...information and (e) enterprise providing service to said particular patient and includes the step of, **deleting redundant duplicate requested information**.

14 A method according to claim 1, including the step of said merging step includes...

...in

response to user command.

28

. A system for use in a healthcare system for **consolidating** patient related

information from **multiple** different sources, comprising: an interface processor for, receiving patient identification data identifying a particular patient related information.

17 A healthcare user interface system supporting user access to **consolidated** patient related **information** derived from **multiple** different sources and including a method comprising the steps of: receiving patient identification data identifying...

00992405

DATA SOURCE PRIVACY SCREENING SYSTEMS AND METHODS

SYSTEMES ET PROCEDES DE VISUALISATION PROTEGEE DE SOURCES DE DONNEES

Patent Applicant/Assignee:

PRIVASOURCE INC, 950 Winter Street, Suite 1700, Waltham, MA 02451, US, US

(Residence), US (Nationality)

Inventor(s):

ERICKSON Lars Carl, 54 Oakwood Road, Newtonville, MA 02460, US,

BREITENSTEIN Agneta, **, US,

PETTINI Donald J, **, US,

Legal Representative:

KELLY Edward J (et al) (agent), Ropes & Gray, One International Place,

Boston, MA 02110-2624, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200321473 A1 20030313 (WO 0321473)

Application: WO 2002US27818 20020830 (PCT/WO US0227818)

Priority Application: US 2001315755 20010830; US 2001315751 20010830; US

2001315753 20010830; US 2001315754 20010830; US 2001335787 20011205

Designated States: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU

CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP

KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO

RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG UZ VN YU ZA ZM ZW

(EP) AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LU MC NL PT SE SK TR

(OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZM ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Publication Language: English

Filing Language: English

Fulltext Word Count: 6713

Main International Patent Class: **G06F-017/00**

Fulltext Availability:

Detailed Description

Claims

English Abstract

...the input dataset could be re-identified from that output dataset.

Individual entities may include **patients** in a **hospital** or served by an insurance carrier, as well as voters, subscribers, customers, companies, or any...

...data acquisition equipment, such as a biologic data sampling device, to prevent de-identification of **patient** or other confidential data acquired by the equipment.

French Abstract

...reidentifiee a partir dudit ensemble de donnees de sortie. Lesdites entites individuelles peuvent etre les **patients** d'un hopital ou des personnes beneficiant des services d'une societe d'assurance, ainsi...

...un dispositif d'echantillonnage de donnees biologiques, afin d'empêcher la desidentification des donnees des **patients** ou d'autres donnees confidentielles acquises par l'equipement.

Detailed Description

... first record to said second fields and said corresponding second values of all of said **second records**; and based on said **comparing**, extracting said first records and said first corresponding values of the highest priority first fields...3 52% 48%

Total 33% 67%

As mentioned above, there were two entries for the **same person** (**identifier** #6). **Records** with **multiple** occurrences belonging to a single person can be more easily identifiable. Consequently, not just the ...time, as well as a count of the number of occurrences of each unique record **identified** within that stream of **data** over the **same** time period.

1 5 Also included is the capacity to tally various record identifiers, such...

...the scrubbed data no longer identifying.

For example, a data clearinghouse may buy personal claims data from multiple insurance companies and sell the combined data to pharmaceutical companies for marketing research. Regulations require that the data be de-identified...

Claim

... first record to said second fields and said corresponding second values of all of said second records ; and based on said comparing , extracting said first records and said first corresponding values of the highest priority first fields...

14/3,K/24 (Item 15 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2004 WIPO/Univentio. All rts. reserv.

00485871 **Image available**

APROBABILISTIC SYSTEM FOR NATURAL LANGUAGE PROCESSING SYSTEME DE PROBABILITES POUR TRAITEMENT DU LANGAGE NATUREL

Patent Applicant/Assignee:

IHC HEALTH SERVICES INC,

Inventor(s):

HAUG Peter J,

KOEHLER Spencer B,

CHRISTENSEN Lee M,

GUNDERSEN Michael L,

VAN BREE Rudy E,

Patent and Priority Information (Country, Number, Date):

Patent: WO 9917223 A1 19990408

Application: WO 98US20463 19980930 (PCT/WO US9820463)

Priority Application: US 9760567 19970930

Designated States: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES

FI GB GE GH GM HR HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD

MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG UZ

VN YU ZW GH GM KE LS MW SD SZ UG ZW AM AZ BY KG KZ MD RU TJ TM AT BE CH

CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE BF BJ CF CG CI CM GA GN GW

ML MR NE SN TD TG

Publication Language: English

Fulltext Word Count: 6758

Main International Patent Class: G06F-017/27

International Patent Class: G06F-017/21

Fulltext Availability:

Detailed Description

English Abstract

...language understanding is described which provides for the generation of concept codes from free-text medical data. A probabilistic model of lexical semantics, in the preferred embodiment of the invention implemented...

Detailed Description

... quality assurance, analyses done for management purposes and in a variety of focused reports that combine information from multiple sources, but is not readily accessible to a human reader unless translation of the coded...extracting its relevant information, oracle text in" calls "admission-text-out" which exists in the same file . That routine checks that the patient is of the right type, i.e., "inpatient". It then calls the...

?

?ds

Set	Items	Description
S1	4328503	DUPLICAT? OR DUPE? ? OR COMMON? OR SAME? OR REDUNDAN? OR R- EPLICA? OR REPEAT? OR IDENTICAL? OR DOUBLE? OR ANALAGOUS? OR - SECOND OR PARALLEL?
S2	2897100	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION - OR CITATION?.
S3	2897155	RECORD OR RECORDS OR DATA OR FILE OR FILES OR INFORMATION - OR CITATION?
S4	1851354	REMOV? OR DELETE? OR DELETING? OR ERASE? ? OR ERASING? OR - DEDUPE? ? OR CHECK? OR FIND????
S5	129201	S1(3N)S2
S6	129206	S1(3N)S3
S7	1995	S6 (3N)S4
S8	2139702	MERGE? OR MERGING? OR CONSOLIDAT? OR COMBIN? OR JOIN??? OR GROUP?
S9	2757	S8 (3N)S6
S10	45	S9 AND S7
S11	442130	MEDICAL? OR PATIENT? OR HEALTH? OR HOSPITAL?
S12	0	S11 AND S10
S13	2	S10 AND IC=(G06F-007/00 OR G06F-017/30)
S14	26	S7 AND S11
S15	2	S14 AND IC=(G06F-007/00 OR G06F-017/30)
S16	80	S7 AND IC=(G06F-007/00 OR G06F-017/30)
S17	7	S16 AND MC=(S05-G02G OR T01-J05B4M OR T01-J06A1)

?show files

File 347:JAPIO Nov 1976-2003/Nov(Updated 040308)

(c) 2004 JPO & JAPIO

File 350:Derwent WPIX 1963-2004/UD,UM &UP=200419

(c) 2004 Thomson Derwent

?

13/5/1 (Item 1 from file: 347)
DIALOG(R) File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

06795294 **Image available**
INFORMATION RETRIEVAL DEVICE, INFORMATION COMPRESSING METHOD FOR
INFORMATION RETRIEVAL DEVICE, AND RECORDING MEDIUM

PUB. NO.: 2001-022775 [JP 2001022775 A]
PUBLISHED: January 26, 2001 (20010126)
INVENTOR(s): ITO MASAO
APPLICANT(s): MATSUSHITA ELECTRIC IND CO LTD
APPL. NO.: 11-194740 [JP 99194740]
FILED: July 08, 1999 (19990708)
INTL CLASS: G06F-017/30 ; G06F-005/00; H03M-007/30

ABSTRACT

PROBLEM TO BE SOLVED: To suppress the capacity of an index file by compressing index information more efficiently and to enable high-speed retrieval using the index file.

SOLUTION: When entries of index **information** are successively **identical**, this **information** compressing method **deletes** by an index information compression part 107 the entries except only one entry, puts (n) successive pieces of index information with the same document number in one **group** as to index **information** in the **same** character chain, and generates compressed index information having the document number compressed into one. When byte data of position numbers are not larger than a reference value (0×0f), the low-order four bits of each of the byte data of the position numbers are connected to generate compressed index information having the position number converted into 1-byte data.

COPYRIGHT: (C)2001,JPO

13/5/2 (Item 2 from file: 347)
DIALOG(R) File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

06061756 **Image available**
DUPLICATED DATABASE MATCHING DEVICE AND METHOD

PUB. NO.: 11-003263 [JP 11003263 A]
PUBLISHED: January 06, 1999 (19990106)
INVENTOR(s): OKAMOTO TOMOYUKI
APPLICANT(s): NEC SOFTWARE LTD
APPL. NO.: 09-154920 [JP 97154920]
FILED: June 12, 1997 (19970612)
INTL CLASS: G06F-012/00; G06F-017/30

ABSTRACT

PROBLEM TO BE SOLVED: To assure the consistency of data despite their updating by **merging** a **duplicated** updating history **information file** and **checking** the consistency of reflection information stored in the file.

SOLUTION: When an application 122 updates a database 111, a DBMS 110 outputs the updating information 113 to an updating history information file 112 for recovery after the base 111 is updated. A reflection information duplicating device 101 produces the reflection information 106 from the information 113 and stores the information 106 in a reflection information file 104. The device 101 transfers the information 106 to a reflection information duplicating device 102 as designated. The device 102 of a duplicating destination database system 121 receives the transferred information 106 and stores it in a duplicated updating history information

file 105. Then a reflection information reflecting device 103 checks the consistency of the reflection information 107 stored in the file 105 and acquires the information 107 to updating it by reflecting it on a database 114.

COPYRIGHT: (C)1999,JPO

?

?t s17/5/1-7

17/5/1 (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

015988062

WPI Acc No: 2004-145912/200415

XRAM Acc No: C04-058844

XRPX Acc No: N04-116280

Establishing or modifying data source comprising entries related to sequences, comprises adding information for retrieving information from a data set to entry in reference data source related to a hash key generated from the sequence

Patent Assignee: LION BIOSCIENCE AG (LION-N)

Inventor: OHR C

Number of Countries: 102 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
EP 1387292	A1	20040204	EP 200216796	A	20020726	200415 B
WO 200413769	A2	20040212	WO 2003EP7743	A	20030716	200415

Priority Applications (No Type Date): EP 200216796 A 20020726

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
-----------	------	-----	----	----------	--------------

EP 1387292	A1	E	20	G06F-017/30	
------------	----	---	----	-------------	--

Designated States (Regional): AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI SK TR

WO 200413769	A2	E		G06F-017/30	
--------------	----	---	--	-------------	--

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SC SD SE SG SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM ZW

Designated States (Regional): AT BE BG CH CY CZ DE DK EA EE ES FI FR GB GH GM GR HU IE IT KE LS LU MC MW MZ NL OA PT RO SD SE SI SK SL SZ TR TZ UG ZM ZW

Abstract (Basic): EP 1387292 A1

NOVELTY - Establishing or modifying a data source comprising entries related to sequences, comprises retrieving a sequence contained in the data set and generating a hash key by applying collision-free hash function; and for each data set, adding information for retrieving information from the data set to entry in reference data source related to a hash key generated from a sequence in the data set.

DETAILED DESCRIPTION - Establishing or modifying a data source comprising entries related to biological sequences that are non-redundant with regard to the sequences on the basis of data sets of one or more basic data sources, each of data sets comprising biological sequence, comprises retrieving for one or more data sets a biological sequence contained in the data set and generating a hash key from the biological sequence thus retrieved by applying a collision-free hash function. The hash function maps the data representing the sequence onto a message of a length shorter than the length of the original data representing the sequence. For each of the data sets, information is added for retrieving information from the data set to an entry in a reference data source uniquely related to the hash key generated from the sequence contained in the data set. A new entry in the reference data source is provided which comprises one unique hash key and information for retrieving the data set or data sets comprising the sequence from which the hash key was generated, if the reference data source does not comprise an entry related to the hash key, such that each entry in the reference data source is uniquely identified by a hash key generated from a sequence.

INDEPENDENT CLAIMS are also included for:

(a) a computer system comprising a reference data source and/or a non-redundant unified database; software for accessing a database

related to biological sequences, each entry in the database being uniquely related to one biological sequence; and software for accessing a reference data source comprising entries, each uniquely related to a biological sequence and having as unique identifier a hash key created by a collision-free hash function mapping the sequence on a data string of shorter length; and

(b) a method of comparing a non-redundant data source comprising entries, each entry being uniquely related to a biological sequence, to data sets of one of more basic data sources related to the sequences, comprising providing a first reference data source comprising entries, each corresponding to exactly one entry in the reference data source, each entry in the reference data source having as unique identifier a hash key obtainable by a collision-free hash function mapping the sequence to which the entry relates on a data string of shorter length; providing a second reference data source comprising entries, each of which is related to exactly one biological sequence and has as a unique identifier the hash key created by the hash function, when applied on the corresponding sequence, where each entry comprises information related to one or more data sets related to the sequence in one of the basic data sources; and comparing the first and second reference data source as to whether entries in the two reference data sources relating to the same hash key comprise different information and/or as to whether an entry in the first or second reference data source is missing in the respective other reference data source.

USE - For establishing or modifying a data source comprising entries related to biological sequences that are non-redundant with regard to the sequences on the basis of data sets of one or more basic data sources, each of data sets comprising biological sequence (claimed).

ADVANTAGE - The method establishes and maintains a non-redundant database with less memory consumption and more algorithmic efficiency, which ideally would also allows for incremental updates.

pp; 20 DwgNo 0/0

Title Terms: ESTABLISH; MODIFIED; DATA; SOURCE; COMPRISE; ENTER; RELATED; SEQUENCE; COMPRISE; ADD; INFORMATION; RETRIEVAL; INFORMATION; DATA; SET; ENTER; REFERENCE; DATA; SOURCE; RELATED; HASH; KEY; GENERATE; SEQUENCE
Derwent Class: B04; D16; S03; T01
International Patent Class (Main): G06F-017/30
International Patent Class (Additional): G06F-019/00
File Segment: CPI; EPI

17/5/2 (Item 2 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

015684075 **Image available**

WPI Acc No: 2003-746264/200370

XRPX Acc No: N03-597946

Data family record managing method, involves adding designated record to family of records when determined that designated record is not duplicate of data records in family, and setting indicator to indicate relationship

Patent Assignee: MEINIG K (MEIN-I)

Inventor: MEINIG K

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20030167253	A1	20030904	US 200291378	A	20020304	200370 B

Priority Applications (No Type Date): US 200291378 A 20020304

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20030167253	A1		39	G06F-007/00	

Abstract (Basic): US 20030167253 A1

NOVELTY - The method involves adding a designated record to the potential family of records when it is automatically determined that

the designated record is not a duplicate of the record in the family. An indicator is automatically set in each of the data records in the potential family of records to indicate a family relationship between the records.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (a) a computer-readable memory medium
- (b) a record management system.

USE - Used for identifying relationship among data records.

ADVANTAGE - The method automatically determines missing information from a particular data **record** and also **deletes duplicate records**

DESCRIPTION OF DRAWING(S) - The drawing shows an example flowchart of a routine for importing and automatically de-duplicating a file of new data records.

pp; 39 DwgNo 7B/24

Title Terms: DATA; FAMILY; RECORD; MANAGE; METHOD; ADD; DESIGNATED; RECORD; FAMILY; RECORD; DETERMINE; DESIGNATED; RECORD; DUPLICATE; DATA; RECORD; FAMILY; SET; INDICATE; INDICATE; RELATED

Derwent Class: T01

International Patent Class (Main): **G06F-007/00**

File Segment: EPI

17/5/3 (Item 3 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

015447549 **Image available**

WPI Acc No: 2003-509691/200348

XPX Acc No: N03-404763

Customer database generation apparatus determines equivalence degree between address and name of customer, based on data equivalence condition of data equivalence determination definition file

Patent Assignee: NIPPON TELEGRAPH & TELEPHONE CORP (NITE)

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 2003173345	A	20030620	JP 2001371635	A	20011205	200348 B

Priority Applications (No Type Date): JP 2001371635 A 20011205

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
JP 2003173345	A		12	G06F-017/30	

Abstract (Basic): JP 2003173345 A

NOVELTY - The selection unit selects the collation data from received address and name information of customer. A collation unit collates the address and name based on the collation rule. The equivalence value of collation result is calculated. A collation unit (15) determines equivalence degree between address and name of customer based on data equivalence conditions of data equivalence determination file.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (1) database generation method; and
- (2) database generation processing program.

USE - For upgrading of databases such as correction of declared mistake, declared unification and **checking of duplicated data**.

ADVANTAGE - Exact collation is enabled even if the information of the address or name is imperfect, **duplicated data is removed** and upgrading of a database is enabled.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of database generation apparatus. (Drawing includes non-English language text).

analysis unit (12,13)
collation unit (15)

pp; 12 DwgNo 1/1
Title Terms: CUSTOMER; DATABASE; GENERATE; APPARATUS; DETERMINE;
EQUIVALENCE; DEGREE; ADDRESS; NAME; CUSTOMER; BASED; DATA; EQUIVALENCE;
CONDITION; DATA; EQUIVALENCE; DETERMINE; DEFINE; FILE
Derwent Class: T01
International Patent Class (Main): G06F-017/30
International Patent Class (Additional): G06F-017/60
File Segment: EPI

17/5/4 (Item 4 from file: 350)
DIALOG(R) File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

015143203 **Image available**
WPI Acc No: 2003-203730/200320
XRPX Acc No: N03-162355

Database information management for duplicate records removal from marketing and business analysis data sets, includes comparing two data sets to detect duplicate records by dividing first data set into U subsets and second into V subsets

Patent Assignee: DATACTICS LTD (DATA-N)

Inventor: BEARDER S; RASCH J

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
GB 2378789	A	20030219	GB 200210164	A	20020503	200320 B

Priority Applications (No Type Date): GB 200111648 A 20010512

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
GB 2378789	A	26	G06F-017/30	

Abstract (Basic): GB 2378789 A

NOVELTY - A method of comparing two data sets, each of which contains a number of records, to detect duplicate records, comprising: dividing the first data set into U subsets; dividing the second dataset into V subsets, at least one of U and V being greater than 1; and comparing each subset, at least some or all of the comparisons being carried out simultaneously in parallel by parallel processors. The subsets of the first and second data sets contain approximately equal number of records.

DETAILED DESCRIPTION - AN INDEPENDENT CLAIM is also included for a data processing system.

USE - Database information management for the removal of duplicate records from large data sets, particularly for client profiling, marketing and business analysis.

ADVANTAGE - Provides a method for removing duplicate records from a data set in a more efficient and economical manner with regard to the costs involved in processor memory and time, while ensuring that performance doesn't depend on the relevance and accuracy of the chosen sorting key and that the comparison is not total.

DESCRIPTION OF DRAWING(S) - The drawing shows a block diagram that illustrates a central processor receiving data sets S and T and divides these into subsets.

pp; 26 DwgNo 1/2

Title Terms: DATABASE; INFORMATION; MANAGEMENT; DUPLICATE; RECORD; REMOVE;
MARKET; BUSINESS; ANALYSE; DATA; SET; COMPARE; TWO; DATA; SET; DETECT;
DUPLICATE; RECORD; DIVIDE; FIRST; DATA; SET; SUBSET; SECOND; SUBSET
Derwent Class: T01
International Patent Class (Main): G06F-017/30
File Segment: EPI

17/5/5 (Item 5 from file: 350)
DIALOG(R) File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

015030892 **Image available**

WPI Acc No: 2003-091409/200308

XRPX Acc No: N03-072344

Database management method involves identifying data cells which satisfies input data filtering criteria, and sorting data sets of identified cells based on priority list for display

Patent Assignee: GORMAN J G (GORM-I)

Inventor: GORMAN J G

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20020143780	A1	20021003	US 2001823406	A	20010330	200308 B

Priority Applications (No Type Date): US 2001823406 A 20010330

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
US 20020143780	A1	25	G06F-007/00	

Abstract (Basic): US 20020143780 A1

NOVELTY - The order of data filtering criteria input by the user is recorded and a sorting priority list defining a sorting priority for data cells in data table is generated. The data cells satisfying the filtering criteria are identified and displayed interleaved between non-identified cells. The data set of identified cells are sorted based on priority list and graphically displayed to the user by highlighting it.

DETAILED DESCRIPTION - An INDEPENDENT CLAIM is included for database management system.

USE - For managing database storing business information, educational information, personal information, medical information and tax data.

ADVANTAGE - The sorting process enables the user to **find** desired **data** on the **same** screen.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of data management system.

pp; 25 DwgNo 2/14

Title Terms: DATABASE; MANAGEMENT; METHOD; IDENTIFY; DATA; CELL; SATISFY; INPUT; DATA; FILTER; CRITERIA; SORT; DATA; SET; IDENTIFY; CELL; BASED; PRIORITY; LIST; DISPLAY

Derwent Class: S05; T01

International Patent Class (Main): **G06F-007/00**

File Segment: EPI

17/5/6 (Item 6 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

014132097

WPI Acc No: 2001-616308/200171

XRPX Acc No: N01-459727

Multi-record data storing process for computers, involves inserting append tag defining common value of data field shared by data array, in append section preceding respective data array

Patent Assignee: INPHARMATICA LTD (INPH-N); MARTIN A C R (MART-I)

Inventor: MARTIN A C R; MARTIN A

Number of Countries: 095 Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 200169413	A2	20010920	WO 2001GB1123	A	20010314	200171 B
AU 200140835	A	20010924	AU 200140835	A	20010314	200208
US 20040030502	A1	20040212	WO 2001GB1123	A	20010314	200412
			US 2003221832	A	20030211	

Priority Applications (No Type Date): GB 20006147 A 20000314

Patent Details:

Patent No Kind Lan Pg in IPC Filing Notes

WO 200169413 A2 E 20 G06F-017/00

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA
CH CN CO CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS
JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL
PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR
IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

AU 200140835 A G06F-017/00 Based on patent WO 200169413

US 20040030502 A1 G06F-007/00

Abstract (Basic): WO 200169413 A2

NOVELTY - An array of data sharing a common value for a particular data field is selected. An append tag defining the common value of the data field, is inserted in a append section preceding the data array such that each data inherit the common value, when the selected data file is read.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (a) Data file;
- (b) Computer system;
- (c) Computer program product

USE - For storing large array of data generated in digital imaging and database compilation in computers.

ADVANTAGE - As the append tag defining common value of data field shared by the **data** array is included, **redundancy** in a **data file** is **removed**, markup of data is minimized, files are parsed and transmitted more quickly, less storage memory is required, additional information can be added without designing special parser.

pp; 20 DwgNo 0/0

Title Terms: MULTI; RECORD; DATA; STORAGE; PROCESS; COMPUTER; INSERT;
APPENDAGE; TAG; DEFINE; COMMON; VALUE; DATA; FIELD; SHARE; DATA; ARRAY;
APPENDAGE; SECTION; PRECEDE; RESPECTIVE; DATA; ARRAY

Derwent Class: T01

International Patent Class (Main): G06F-007/00 ; G06F-017/00

International Patent Class (Additional): G01N-033/48; G01N-033/50;

G06F-017/30 ; G06F-019/00

File Segment: EPI

17/5/7 (Item 7 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

011145597 **Image available**

WPI Acc No: 1997-123521/199712

XRPX Acc No: N97-101810

Correlation checking method for common data in online processing system - involves performing simultaneous updation of common data stored by each section, if correlation check result is judged to be normal

Patent Assignee: HITACHI LTD (HITA)

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 9006660	A	19970110	JP 95154285	A	19950621	199712 B

Priority Applications (No Type Date): JP 95154285 A 19950621

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes

JP 9006660 A 6 G06F-012/00

Abstract (Basic): JP 9006660 A

The method involves judging need for updation of each section, when data is inserted into each section. When updation is needed, correlation check between these sections is performed.

When the **check** result is normal, **common data** stored by each section is updated simultaneously.

ADVANTAGE - Restores duplication of data insertion in relative sections. Improves precision of database between each section.

Dwg.5/5

Title Terms: CORRELATE; CHECK; METHOD; COMMON; DATA; PROCESS; SYSTEM;
PERFORMANCE; SIMULTANEOUS; COMMON; DATA; STORAGE; SECTION; CORRELATE;
CHECK; RESULT; JUDGEMENT; NORMAL

Derwent Class: T01

International Patent Class (Main): G06F-012/00

International Patent Class (Additional): G06F-015/00; **G06F-017/30**

File Segment: EPI

?

	Hits	Search Text	DBs	Time Stamp Δ
1	12019	707/1-5,8-10.ccls.	USPAT; US-PGPUB	2004/05/20 09:24
2	6571	707/100,102,104.1.ccls.	USPAT; US-PGPUB	2004/05/20 09:24
3	4263	707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/20 09:25
4	11224	709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/21 09:07
5	1171	705/1-3.ccls.	USPAT	2004/05/21 09:07
6	11224	709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/21 09:07
7	18581	707/1-5,8-10.ccls. or 707/100,102,104.1.ccls. or 707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/21 09:08
8	18581	707/1-5,8-10.ccls. or 707/100,102,104.1.ccls. or 707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/21 10:36
9	1	5497486.pn.	USPAT	2004/05/21 10:37
10	1	5497486.pn.	USPAT	2004/05/21 10:37
11	1	6480859.pn.	USPAT	2004/05/21 10:38
12	8	(determin\$4 with commonality) and likelihood and record\$2	USPAT	2004/05/21 10:52
13	9	((determin\$4 or establish\$4 or defin\$4) with commonality) and likelihood and record\$2	USPAT	2004/05/21 10:56
14	83	(identif\$5 adj2 record\$2) and ((appl\$5 or us\$4) and (record\$2 with match\$5) and criteria) and (compar\$4 with element\$2)	USPAT	2004/05/21 10:57
15	18581	707/1-5,8-10.ccls. or 707/100,102,104.1.ccls. or 707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/23 13:13
16	15359	705/1-3.ccls. or 709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/23 13:14
17	15359	705/1-3.ccls. or 709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/23 13:14
18	18	(determin\$4 with commonality) and likelihood	USPAT	2004/05/23 13:19
19	1006	(determin\$4 with (commonality or agreement)) and (likelihood or probability or possibility)	USPAT	2004/05/23 13:20
20	5121	identif\$5 adj2 record\$2	USPAT	2004/05/23 13:20
21	2459	(appl\$5 or us\$4) and (record\$2 with match\$5) and criteria	USPAT	2004/05/23 13:21
22	69765	compar\$4 with element\$2	USPAT	2004/05/23 13:22

	Hits	Search Text	DBs	Time Stamp ^Δ
23	69765	compar\$4 with element\$2	USPAT	2004/05/23 13:22
24	36	((determin\$4 with (commonality or agreement)) and (likelihood or probability or possibility)) and (identif\$5 adj2 record\$2)	USPAT	2004/05/23 13:23
25	26	((determin\$4 with (commonality or agreement)) and (likelihood or probability or possibility)) and (identif\$5 adj2 record\$2) and ((appl\$5 or us\$4) and (record\$2 with match\$5) and criteria)	USPAT	2004/05/23 13:23
26	3	((determin\$4 with (commonality or agreement)) and (likelihood or probability or possibility)) and (identif\$5 adj2 record\$2) and ((appl\$5 or us\$4) and (record\$2 with match\$5) and criteria) and (compar\$4 with element\$2)	USPAT	2004/05/23 13:23
27	18581	707/1-5,8-10.ccls. or 707/100,102,104.1.ccls. or 707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/24 07:51
28	15359	705/1-3.ccls. or 709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/24 07:51
29	18621	707/1-5,8-10.ccls. or 707/100,102,104.1.ccls. or 707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/25 13:17
30	15385	705/1-3.ccls. or 709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/25 13:17
31	18621	707/1-5,8-10.ccls. or 707/100,102,104.1.ccls. or 707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/26 07:41
32	15385	705/1-3.ccls. or 709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/26 07:41
33	5132	identif\$5 adj2 record\$2	USPAT	2004/05/26 08:01
34	7462	electro\$2cardiograph or ECG	USPAT	2004/05/26 08:05
35	2279	electro\$2encepholograph or EEG	USPAT	2004/05/26 08:06
36	101787	x\$2ray	USPAT	2004/05/26 08:07
37	10247	laboratory adj2 test\$2 and result\$2	USPAT	2004/05/26 08:08
38	46772	physical adj2 characteristic\$2	USPAT	2004/05/26 08:09
39	16736	previous and diagnosis	USPAT	2004/05/26 08:11

	Hits	Search Text	DBs	Time Stamp ^Δ
40	38797	ventilation	USPAT	2004/05/26 08:12
41	3535	infusion adj2 pump\$2	USPAT	2004/05/26 08:14
42	70499	patient\$2 and clinical	USPAT	2004/05/26 08:16
43	859	medical adj2 services	USPAT	2004/05/26 08:17
44	281	"non-clinical" and data	USPAT	2004/05/26 08:17
45	2469	(appl\$5 or us\$4) and (record\$2 with match\$5) and criteria	USPAT	2004/05/26 08:20
46	3	(identif\$5 adj2 record\$2) and ((appl\$5 or us\$4) and (record\$2 with match\$5) and criteria) and ((electro\$2cardiograph or ECG) or (electro\$2encephalograph or EEG)) and (patient\$2 and clinical)	USPAT	2004/05/26 08:33
47	18681	707/1-5,8-10.ccls. or 707/100,102,104.1.ccls. or 707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/27 07:22
48	15456	705/1-3.ccls. or 709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/27 07:22
49	1	(identif\$5 adj2 record\$2) and ((appl\$5 or us\$4) and (record\$2 with match\$5) and criteria) and (laboratory adj2 test\$2 and result\$2) and (physical adj2 characteristic\$2) and (patient\$2 and clinical)	USPAT	2004/05/27 11:35
50	7	(identif\$5 adj2 record\$2) and ((appl\$5 or us\$4) and (record\$2 with match\$5) and criteria) and (laboratory adj2 test\$2 and result\$2) and (patient\$2 and clinical)	USPAT	2004/05/27 11:36
51	18581	707/1-5,8-10.ccls. or 707/100,102,104.1.ccls. or 707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/27 13:46
52	18681	707/1-5,8-10.ccls. or 707/100,102,104.1.ccls. or 707/200-203,205,206.ccls.	USPAT; US-PGPUB	2004/05/27 13:46
53	12350	705/1-3.ccls. or 709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/27 13:47
54	5121	identif\$5 adj2 record\$2	USPAT	2004/05/27 13:47
55	2459	(appl\$5 or us\$4) and (record\$2 with match\$5) and criteria	USPAT	2004/05/27 13:47

	Hits	Search Text	DBs	Time Stamp ^Δ
56	15456	705/1-3.ccls. or 709/200-203,223,224.ccls.	USPAT; US-PGPUB	2004/05/27 13:47
57	5132	identif\$5 adj2 record\$2	USPAT	2004/05/27 13:47
58	28834	allerg\$4	USPAT	2004/05/27 13:48
59	426	ventilation with data	USPAT	2004/05/27 13:48
60	2469	(appl\$5 or us\$4) and (record\$2 with match\$5) and criteria	USPAT	2004/05/27 13:48
61	28834	allerg\$4	USPAT	2004/05/27 13:48
62	426	ventilation with data	USPAT	2004/05/27 13:48
63	9477	blood with oxygen	USPAT	2004/05/27 13:49
64	39210	blood with (oxygen or pressure)	USPAT	2004/05/27 13:49
65	762	infusion adj2 pump\$2 and pulse	USPAT	2004/05/27 13:50
66	762	infusion adj2 pump\$2 and pulse	USPAT	2004/05/27 14:04
67	3	(identif\$5 adj2 record\$2) and ((appl\$5 or us\$4) and (record\$2 with match\$5) and criteria) and allerg\$4 and (blood with (oxygen or pressure))	USPAT	2004/05/27 14:05
68	3535	infusion adj2 pump\$2	USPAT	2004/05/27 14:05
69	1	(identif\$5 adj2 record\$2) and ((appl\$5 or us\$4) and (record\$2 with match\$5) and criteria) and (ventilation with data)	USPAT	2004/05/27 14:49
70	6	(identif\$5 adj2 record\$2) and ((appl\$5 or us\$4) and (record\$2 with match\$5) and criteria) and ((electro\$2cardiograph or ECG) or (electro\$2encepholograph or EEG))	USPAT	2004/05/27 14:50
71	4	(identif\$5 adj2 record\$2) and (infusion adj2 pump\$2 and pulse)	USPAT	2004/05/27 14:53